

UNIVERSIDAD DE COSTA RICA  
SISTEMA DE ESTUDIOS DE POSGRADO

CARACTERIZACIÓN DE DIVERSIDAD GENÉTICA EN POBLACIONES  
CULTIVADAS DE CAS (*Psidium friedrichsthalianum* (O. Berg) Nied) UTILIZANDO  
DESCRIPTORES CITOGENÉTICOS Y MOLECULARES PARA FINES DE  
MEJORAMIENTO GENÉTICO

Tesis sometida a la consideración de la Comisión del Programa de Estudios de Posgrado  
en Biología para optar al grado y título de Maestría académica en Biología con Énfasis  
en Genética y Biología Molecular

MÓNICA MARÍA ROJAS GÓMEZ

Ciudad Universitaria Rodrigo Facio, Costa Rica

2020

## **Dedicatoria**

A mi familia: Magaly Gómez, Roberto Rojas y Francisco García.

## **Agradecimientos**

Mi principal agradecimiento es hacia Dios y mi familia por apoyarme en todo el proceso de la maestría y motivarme siempre cuando lo necesitaba.

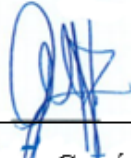
Agradezco la ayuda de la Estación Experimental Agrícola Fabio Baudrit Moreno (Universidad de Costa Rica) principalmente a Pablo Bolaños por la capacitación en la tinción de cromosomas y en el uso del microscopio de fluorescencia; agradezco también al Centro Nacional de Innovaciones Biotecnológicas (CENIBiot) en CeNAT-CONARE y a la Finca Los Diamantes en el Instituto Nacional de Innovación y Transferencia en Tecnología Agropecuaria (INTA). De igual manera, agradezco a la Vicerrectoría de Investigación de la Universidad de Costa Rica (UCR) por subvencionar una parte del proyecto (subvención 111-B7-26) y al Consejo Nacional de Rectores (CONARE) por la subvención del FEES-15-2019, al cual pertenece este proyecto.

Por otra parte, agradezco al Centro de Investigación en Biología Celular y Molecular de la UCR (CIBCM), por permitirme hacer los análisis en el citómetro de flujo y especialmente un gran agradecimiento a Alfonso García por capacitarme para usar el equipo, así como por la guía y recomendaciones brindadas. De igual manera agradezco enormemente a Laura Brenes del CIBCM por guiarme con los procesos del análisis bioinformático de las secuencias obtenidas por ddRADseq, fue una excelente maestra en el tema.

Así mismo, mi tesis no hubiera sido posible sin la colaboración de los agricultores locales: David Badilla, Gregorio Menocal, Juan Pablo Orozco, Alfonso Ruíz, José Fuentes, los cuales me proporcionaron muestras gratuitas de “Cas” de sus fincas para realizar los análisis, les agradezco a todos ellos.

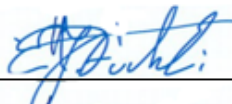
Quiero agradecer enormemente a mi comité de tesis: a mi tutor, Eric Fuchs por el gran apoyo incondicional en todo el proceso de tesis y maestría, siempre estuvo ahí cuando lo necesitaba con buenas ideas y recomendaciones; también a Griselda Arrieta por estar pendiente siempre de los avances de mi tesis y también por el gran apoyo que me brindó con la preparación del examen de candidatura. Agradezco de igual manera a Max Chavarría por estar siempre pendiente de los avances del proyecto en general, así mismo de las recomendaciones brindadas para escribir los manuscritos de mi tesis y consejos para someterlos en las revistas científicas. De los tres he aprendido muchísimo durante este tiempo y su contribución intelectual en este trabajo ha sido invaluable.

“Esta Tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Biología de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Biología con énfasis en Genética y Biología Molecular”



---

Dr. Alfonso García Piñeres  
**Representante del Decano Sistema de Estudios de Posgrado**



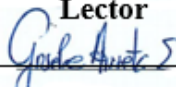
---

Dr. Eric Fuchs Castillo  
**Profesor Guía**



---

Dr. Max Chavarría Vargas  
**Lector**



---

M.Sc. Griselda Arrieta Espinoza  
**Lectora**



---

Dr. Federico Albertazzi Castro  
**Representante del Director del Programa de Posgrado en Biología**



---

Mónica María Rojas Gómez  
**Sustentante**

## Tabla de contenidos

<b>Dedicatoria</b> .....	II
<b>Agradecimientos</b> .....	III
<b>Resumen</b> .....	VII
<b>Abstract</b> .....	VIII
<b>Lista de Cuadros</b> .....	IX
Capítulo 1 .....	IX
Capítulo 2 .....	IX
<b>Lista de Figuras</b> .....	XI
Capítulo 1 .....	XI
Capítulo 2 .....	XI
<b>Material suplementario</b> .....	XII
<b>Lista de Abreviaturas</b> .....	XIII
<b>Introducción general</b> .....	1
Referencias .....	7
<b>CAPÍTULO 1.</b> Genome size and chromosome number of <i>Psidium friedrichsthalianum</i> (O. Berg) Nied (“Cas”) in six populations of Costa Rica .....	14
<b>Abstract</b> .....	16
<b>Introduction</b> .....	17
<b>Materials and Methods</b> .....	19
<i>Sample Collection</i> .....	19
<i>Chromosomal count using DAPI stain</i> .....	19
<i>Flow cytometry estimates of DNA content</i> .....	20
<b>Results and Discussion</b> .....	21
<b>Conclusions</b> .....	25
<b>Acknowledgements</b> .....	25
<b>Disclosure Statement</b> .....	26
<b>Geolocation data</b> .....	26
<b>Funding details</b> .....	26
<b>References</b> .....	26
<b>Tables</b> .....	32
<b>Figures</b> .....	35

<b>CAPÍTULO 2.</b> Caracterización de diversidad y estructura genética de poblaciones cultivadas de “Cas” ( <i>Psidium friedrichsthalianum</i> Ndz) en Costa Rica utilizando ddRADseq <sup>1</sup> .....	39
<b>Resumen</b> .....	40
<b>Introducción</b> .....	41
<b>Materiales y métodos</b> .....	44
<i>Colecta de material vegetal</i> .....	44
<i>Secuenciación con el método ddRAD-seq</i> .....	44
<i>Análisis de secuencias (lecturas) e identificación de SNPs</i> .....	45
<i>Estimativas básicas de diversidad y estructura genética</i> .....	46
<b>Resultados</b> .....	47
<i>Análisis de secuencias (lecturas) e identificación de SNPs</i> .....	47
<i>Estadísticas de diversidad genética</i> .....	47
<i>Estructura genética</i> .....	48
<b>Discusión</b> .....	48
<b>Conclusiones</b> .....	51
<b>Referencias</b> .....	52
CUADROS .....	61
FIGURAS .....	64
<b>MATERIAL SUPLEMENTARIO</b> .....	69
<b>Conclusiones generales</b> .....	74

## Resumen

*Psidium friedrichsthalianum* (O. Berg) Nied es una especie cultivable de origen tropical de la familia Myrtaceae, subfamilia Myrtoideae, tribu Myrteae (Lucas *et al.* 2019). Actualmente, no hay información sobre la cantidad y distribución de la diversidad genética en poblaciones cultivadas de Cas en Costa Rica, la gran mayoría se han realizado en *Psidium guajava* y con marcadores tipo microsatélites (SSR); de igual manera, no se conoce los niveles de ploidía de estas poblaciones, Srivastava (1977) informó que este cultivo es diploide ( $2n = 2x = 11$ ) e Hirano (1967) identificó individuos tetraploides y hexaploides en muestras recolectadas en América Central; es probable que la diversidad en el número cromosómico previamente reportado para Cas puede ser una consecuencia de su proceso de domesticación en curso. La ausencia de dicha información limita el desarrollo de programas de selección y mejoramiento genético en el cultivo de Cas, que sean eventualmente dirigidos a cosechar frutos de Cas con contenido alto y homogéneo de polifenoles, y que sea comercializado como alimento funcional. Por lo tanto, el objetivo del presente estudio fue caracterizar diversidad genética de poblaciones cultivadas de Cas (*Psidium friedrichsthalianum* (O. Berg) Nied) en Costa Rica utilizando descriptores citogenéticos y moleculares para su eventual conservación y/o selección en programas de mejoramiento genético. En el primer capítulo, cuantificamos el número de cromosomas (ploidía) de seis poblaciones cultivadas de Cas mediante tinción de cromosomas con DAPI y determinamos el tamaño del genoma de estas poblaciones utilizando citometría de flujo. Observamos que las poblaciones de Cas en Costa Rica son consistentemente diploides ( $2n = 2x = 22$ ). Así mismo, determinamos que los individuos de dichas poblaciones presentan un contenido promedio de ADN nuclear de  $2C$  de  $1.960 \pm 0.005$  pg, equivalente a 1916.88 Mbp. La uniformidad encontrada entre las poblaciones costarricenses de Cas en términos de número cromosómico y tamaño del genoma (ADN nuclear) sugiere un origen común entre ellas. Alternativamente, sugerimos que dicha uniformidad puede ser una consecuencia del establecimiento de las plantaciones a partir de la selección artificial de individuos con rasgos de interés específicos y homogéneos, por ejemplo, tamaño de la fruta o alto contenido de pulpa. En el segundo capítulo, caracterizamos diversidad y estructura genética en cinco poblaciones cultivadas de Cas en Costa Rica utilizando ddRAD-seq; como resultados obtuvimos 1 148 SNPs lo que sugiere que el método ddRAD-seq fue apto para genotipar dichas poblaciones de Cas. Se obtuvo un valor promedio de heterocigosidad esperada ( $H_e$ ) igual a  $0.352 \pm 0.003$ , lo que sugiere alta diversidad genética en todas las poblaciones analizadas. Así mismo, hubo un exceso de heterocigotas en todas las poblaciones analizadas, posiblemente por causa de establecer estas plantaciones a partir de individuos o variedades altamente heterocigotas. Se realizó un dendrograma tipo Neighbor-Joining (NJ) y un análisis discriminante de componentes principales (DAPC), los cuales mostraron dos grupos de poblaciones, en congruencia con el análisis de STRUCTURE que propone  $K=2$ ; el primer grupo representado por la población de Escazú y el otro grupo distribuido casi homogéneamente entre las demás poblaciones. Nuestros resultados proponen que las poblaciones de Cas se estructuran principalmente por un factor de elevación y barreras naturales, pero no por distancia (IBD). Concluimos que el material genético altamente diverso observado en el segundo capítulo es promisorio para su resguardo o conservación; dicho material puede utilizarse eventualmente para el establecimiento de programas de selección y mejoramiento genético del cultivo de Cas.

## Abstract

*Psidium friedrichsthalianum* (O. Berg) Nied is a cultivable species of tropical origin in the family Myrtaceae, subfamily Myrtoideae, tribe Myrteae (Lucas *et al.* 2019). It produces a fruit popularly known as "Cas", "Acid Guayaba" or "Costa Rican Guayaba" (Barahona and Rivera 1995; Rojas-Rodríguez and Torres-Córdoba 2013) that is highly desired in the diet of Costa Ricans and Central Americans (mainly Nicaragua and El Salvador). Something remarkable about this cultivable species is that its fruit has been potentiated as a "functional food", due to its high polyphenolic content (proanthocyanidins (PACs) type A and B) and its high antioxidant capacity (Red -Garbanzo *et al.* 2019; Cuadrado-Silva *et al.* 2017; Granados-Chinchilla *et al.* 2016; Flores *et al.* 2013; González *et al.* 2012). Currently, there is no information on the can tity and distribution of genetic diversity in cultivated populations of Cas in Costa Rica, the vast majority have been carried out in *Psidium guajava* and with microsatellite-type markers (SSR); likewise, the ploidy levels of these populations are not known, Srivastava (1977) reported that this culture is diploid ( $2n = 2x = 11$ ) and Hirano (1967) identified tetraploid and hexaploid individuals in samples collected in Central America; It is likely that the diversity in chromosome number previously reported for Cas may be a consequence of his ongoing domestication process. Likewise, the size of the genome or C value of Cas is also unknown, it has been characterized only in congeners, for example, *Psidium acutangulum*, *Psidium cattleianum*, *Psidium guajava* L. (white cultivar), *Psidium guajava* L. (red cultivar), *Psidium guineense* and *Psidium grandifolium* (Costa and Forni-Martins 2006b, Costa *et al.* 2008; Machado-Marques *et al.* 2016; Coser *et al.* 2012; Souza *et al.* 2015). In the first chapter, we quantified the chromosome number (ploidy) of six cultured populations of Cas by staining chromosomes with DAPI and determined the genome size of these populations using flow cytometry. We observe that the populations of Cas in Costa Rica are consistently diploid ( $2n = 2x = 22$ ). The results of this first chapter are essential for future breeding programs, hybridization practices and development of QTL (Quantitative Trait Loci). In the second chapter, we characterized diversity and genetic structure in five cultivated populations of Cas in Costa Rica using ddRAD-seq; As results, we obtained 1,148 SNPs, which suggests that the ddRAD-seq method was suitable for genotyping these populations of Cas. An average expected heterozygosity value ( $H_e$ ) equal to  $0.352 \pm 0.003$  was obtained, which suggests high genetic diversity in all the analyzed populations. Likewise, there was an excess of heterozygotes in all the analyzed populations, possibly due to establishing these plantations from highly heterozygous individuals or varieties. On the other hand, we obtained a  $F_{st} = 0.220$  ( $p < 0.05$ ), which evidences high and significant genetic structure among the analyzed populations, in addition, this suggests that the gene flow in these populations is restricted and occurs mainly between nearby or adjacent populations. these. A Neighbor-Joining (NJ) dendrogram and a principal component discriminant analysis (DAPC) were performed, which showed two groups of populations, consistent with the STRUCTURE analysis proposed by  $K = 2$ ; the first group represented by the Escazú population and the other group distributed almost homogeneously among the other populations. Our results propose that Cas populations are structured mainly by an elevation factor and natural barriers, but not by distance (IBD). We conclude that the highly diverse genetic material observed in the second chapter is promising for safekeeping or conservation; Such material may eventually be used for the establishment of breeding and breeding programs for the Cas crop.



## Lista de Cuadros

### Capítulo 1

**Table 1.** Chromosomes number and genome size from different species of *Psidium* and *Eucalyptus* (Myrtaceae) determined in previous studies. The content of holoploid nuclear DNA (2C, pg DNA) and the content of monoploid DNA (1C, pg DNA) are also provided..... 32

**Table 2.** Collection sites of “Costa Rican Guava” used to determine genome size and chromosomal number. **PPT**: mean annual precipitation (mm); **Samples CMF**: number of seedlings used for flow cytometry; **Samples NM**: number of samples used to determine chromosomes count..... 33

**Table 3.** Parameters obtained by flow cytometry to determine the genome size of *Psidium friedrichsthalianum*. **NE**: Number of events obtained; **CV**: Coefficient of variation obtained; **2C (pg)**: holoploid nuclear DNA content obtained; 1pg DNA = 978 Mbp (Dolezel *et al.* 2003; Bennett *et al.* 2000).....34

### Capítulo 2

**Cuadro 1.** Sitios de colecta de material vegetal de “guayaba costarricense” (*P. friedrichsthalianum*) para determinar diversidad y estructura genética. **Población**: nombre de la población colectada, **PMA**: precipitación media anual (mm), **Temp**: temperatura anual. **Región**: zona geográficas de Costa Rica en donde se encuentran ubicadas las poblaciones.....61

**Cuadro 2.** Estimativas básicas de diversidad genética de poblaciones de “guayaba costarricense” (*P. friedrichsthalianum*) obtenidas en el presente estudio. **PIC**= Contenido de información polimórfica, **Ho**= heterocigosidad observada, **He**= heterocigosidad esperada, **Ae**= número efectivo por locus, **Fis**=coeficiente de endogamia, **Fst**= estadístico de estructura genética.....62

<b>Cuadro 3.</b> Distancias genéticas de Nei entre las poblaciones de “guayaba costarricense” ( <i>P. friedrichsthalianum</i> ) analizadas en el presente estudio.....	63
---	----

## Lista de Figuras

### Capítulo 1

**Figure 1.** (a) Seedlings of *P. friedrichsthalianum* used to measure genome size by flow cytometry. (b and c) Flower buds of *P. friedrichsthalianum* used for cytogenetic observations.....36

**Figure 2.** Bivalent chromosomes of *P. friedrichsthalianum* in meiotic metaphase, stained with DAPI, scale bar 10um. (a, b and c) Samples from the populations of Cervantes, Tacacorí and Escazú respectively, showing 11 bivalent chromosomes. (d) Image showing chromosomes in trivalent, bivalent and univalent states.....37

**Figure 3.** Relative fluorescence intensity (propidium iodide (PI)) histogram obtained after a simultaneous cytometric analysis of nuclei of reference standard (*Glycine max*, 2C=2.50 pg of DNA) and *Psidium friedrichsthalianum* (2C= 1.960 ± 0.005 pg) .....38

### Capítulo 2

**Figura 1.** Mapa de ubicaciones geográficas donde se colectó el material vegetal de “guayaba costarricense” (*P. friedrichsthalianum*). Los puntos marcados son puntos exactos del muestreo.....64

**Figura 2.** Pipeline bioinformático desarrollado para el análisis de las lecturas obtenidas por ddRAD-seq. .... 65

**Figura 3.** Dendrograma basado en las distancias genéticas de Nei utilizando el algoritmo de Neighbor-Joining (NJ) entre poblaciones de “guayaba costarricense” (*P. friedrichsthalianum*) .....66

**Figura 4.** Resultados del análisis discriminante de componentes principales (DAPC) basado en los datos de SNPs de “guayaba costarricense”. Los ejes representan los dos primeros discriminadores lineales (LD); cada círculo representa un grupo y cada punto representa un individuo. Los números romanos y eclipses (de color azul) representan las diferentes subpoblaciones identificadas por el análisis DAPC.....67

**Figura 5.** Estructura de cinco poblaciones de “guayaba costarricense” revelada por análisis bayesiano (STRUCTURE), mostrando el mejor conglomerado de K=2. Cada columna vertical representa un individuo y diferentes colores en el mismo individuo indican el porcentaje del genoma compartido con cada grupo, de acuerdo con las proporciones de la mezcla. El eje y denota la probabilidad de la asignación de un individuo en las poblaciones muestreadas.....68

## Material suplementario

**Figura S1.** Imágenes representativas de algunas plantaciones de “guayaba costarricense” en donde se realizó la colecta del material vegetal para el presente estudio.....69

**Figura S2.** Alineamiento visualizado en IGV 2.3.x release de secuencias de algunos individuos de “guayaba costarricense” contra el congénere *Psidium guajava*.....70

**Figura S3.** Gráfica mostrando el número de lecturas antes (columnas azules) y después (columnas naranjas) del procesamiento con Trimmomatic-0.36 (SLIDINGWINDOW:4:30 y MINLEN:100) para cada individuo de las poblaciones de “guayaba costarricense” (*P. friedrichsthalianum*) .....71

**Figura S4.** Porcentaje de transiciones (C↔T o A ↔G) (57.46%) y transversiones (A↔C, C↔G, A↔T, G↔T) (42.54%) observadas en el presente estudio.....72

**Figura S5.** Correlación de la distancia geográfica (Dgeo) y la distancia genética (Dgen) entre 75 individuos de poblaciones de “guayaba costarricense” (*P. friedrichsthalianum*), incluida la línea de regresión (prueba de Mantel, valor de correlación = 0.453, p-value = 0.097, en la aleatorización 999) .....73

## **Lista de Abreviaturas**

RF	Recursos fitogenéticos
A	Adenina
T	Timina
DAPI	4',6-Diamidine-2'-phenylindole dihydrochloride
PI	Propidium iodide
UV	Ultravioleta
NGS	Secuenciación de Nueva Generación
RAD-seq	Secuenciación de ADN asociada a sitios de restricción
ddRAD-seq	Secuenciación de ADN asociada a doble sitio restricción
ADN	Ácido desoxirribonucleico
SNP	Polimorfismo de un solo nucleótido
PACs	Proantocianidinas
QTL	Quantitative Trait Loci
DTT	Dithiothreitol
MFI	Mean fluorescence intensity
FCS	Express 4 Flow Cytometry Software
Mbp	Megabase pairs
ANOVA	Analysis of Variance
TE	Transposable elements
2C	Contenido de ADN nuclear
He	Heterocigosidad esperada
HWE	Equilibrio de Hardy-Weinberg
HSAECs	Human Primary Small Airway Epithelial Cells
CTAB	Cetyl Trimethyl Ammonium Bromide
MAF	Minor Allele Frequency
Ho	Heterocigosidad observada

Ae	Número efectivo por locus
PIC	Contenido de Información Polimórfica
FIS	Coeficiente de endogamia
NJ	Neighbor-Joining
DAPC	Análisis discriminante de componentes principales
MCMC	Monte Carlo de la Cadena de Markov
SSR	Microsatélites



UNIVERSIDAD DE  
COSTA RICA

SEP Sistema de  
Estudios de Posgrado

**Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.**

Yo, Mónica Rojas Gómez, con cédula de identidad 402070943 en mi condición de autor del TFG titulado Caracterización de diversidad genética de poblaciones cultivadas de Cas (*Psidium friedrichthalianum*) utilizando descriptores citogenéticos y moleculares para fines de mejoramiento genético.

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. SI ☒ NO ☐

\*En caso de la negativa favor indicar el tiempo de restricción: \_\_\_\_\_ año (s).

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

**INFORMACIÓN DEL ESTUDIANTE:**

Nombre Completo: Mónica Rojas Gómez

Número de Carné: A85564 Número de cédula: 402070943

Correo Electrónico: morogo27@gmail.com

Fecha: 28/08/20 Número de teléfono: 8874 4043

Nombre del Director (a) de Tesis o Tutor (a): Dr. Eric Fuchs Castillo

Mónica VCG

**FIRMA ESTUDIANTE**

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

## Introducción general

Estudiar la cantidad y distribución de la diversidad genética dentro de una especie cultivable es importante para establecer prácticas eficientes de conservación y reproducción (Awise 1994; Chaveerach *et al.* 2007; El-Esawi 2012); por ejemplo, permite identificar material genético promisorio que puede ser utilizado eventualmente para reproducir nuevos cultivares, favorecer la domesticación o bien desarrollar nuevas variedades según las necesidades productivas de esa especie cultivable (Olsen *et al.* 2013) (Bretting and Widrlechner 1995).

Los descriptores citogenéticos poseen un papel importante en el conocimiento y explotación de la diversidad genética vegetal; ya que parte del principio de que las especies se caracterizan por un número estable de cromosomas y que pueden diferenciarse del resto de las especies o variedades dependiendo de su número cromosómico (ploidía) (Wendel *et al.* 2016). Conocer el número de cromosomas o ploidía de una especie cultivable es esencial para el diseño efectivo de estrategias de mejoramiento genético; permite prácticas de hibridación, desarrollo de QTL (loci de rasgos cuantitativos) y permite comprender mejor los efectos de la endogamia y la heterosis (Birchler 2013; Washburn and Birchler 2014).

Estimar el tamaño del genoma o valor C de una especie cultivable es de gran importancia como recurso fitogenético, ya que puede utilizarse como herramienta para beneficiar las prácticas de mejora genética (Birchler 2013; Washburn and Birchler 2014). El tamaño del genoma o valor C de una especie vegetal corresponde a la cantidad de ADN en su núcleo haploide o gamético no replicado (polen) (Greilhuber *et al.* 2005). Las variaciones del tamaño del genoma entre especies vegetales pueden tener un significado de evolución adaptativa; existen evidencias de correlación entre el tamaño del genoma y diversos rasgos fenotípicos de aparente importancia selectiva en especies vegetales, como



el tamaño de semilla (Wakamiya 1993), respuesta al CO<sub>2</sub> (Jasienski and Bazzaz 1995), eficiencia de tasas metabólicas (Vinogradov 1995), desarrollo de plántulas (Smarda *et al.* 2008), tamaño de la flor (Meagher *et al.* 1994), etc. En cuanto a los caracteres ambientales, Knight y Ackerly (2002) encontraron correlación entre el tamaño del genoma y temperaturas extremas o precipitaciones anuales; así mismo, Achigan-Dako *et al.* (2008) observaron una correlación entre tamaño del genoma y altitud para el cultivo de calabaza.

Una técnica ampliamente utilizada para estimar el tamaño del genoma de una especie vegetal es la citometría de flujo (Doležel and Bartoš 2005). El principio de citometría de flujo consiste en el uso de un agente que es intercalante en los nucleótidos del ADN (principalmente adenina (A) y timina (T)) y que es fluorescente (PI (propidium iodide)), posteriormente, se utiliza luz ultravioleta (UV) o un “Green Laser” para realizar conteos de células o núcleos que van a emitir la fluorescencia previamente añadida y con este conteo se determina el contenido de ADN o tamaño del genoma nuclear (Laguado 2007). Así mismo, con el desarrollo actual de las técnicas de secuenciación de segunda y tercera generación (NGS), la información sobre el tamaño del genoma o el valor C es esencial para establecer condiciones experimentales apropiadas, para preparar de manera efectiva bibliotecas genómicas y secuenciar genomas completos de especies cultivables (Leitch and Leitch 2008).

De la misma manera, los marcadores moleculares son descriptores ampliamente utilizados en estudios de diversidad genética en plantas (Kumar and Agrawal 2017; Treuren and Hintum 2014; Bordoni and Hodgkin 2012). A través de los marcadores moleculares es posible analizar en profundidad la estructura del genoma de las plantas y de esta manera hacer más eficiente la utilización de los recursos genéticos en el mejoramiento vegetal; incluso en los últimos años se ha avanzado mucho en el desarrollo

de mapas genéticos y en el marcado de genes de interés agronómico (Marker Assisted Selection” (MAS)). Un ejemplo de marcadores moleculares son los polimorfismos de un solo nucleótido (SNPs) (Thavamanikumar *et al.* 2011), los SNPs se caracterizan por ser abundantes en el genoma en comparación a otros marcadores, además una gran mayoría de SNPs se encuentran en regiones genómicas codificantes, tienen menor tasa de error y mayor rentabilidad en comparación con los microsatélites (SSR) y AFLPs (Mammadov *et al.* 2012; Kumar *et al.* 2017). Así mismo, los SNPs poseen una baja tasa de mutación ( $\sim 10^{-8}$ ), siendo más adecuados para marcadores de selección a largo plazo (McKay *et al.* 2008; Lin *et al.* 2010).

Con la aparición de la secuenciación de próxima generación (NGS, con sus siglas en inglés) se han implementado diferentes métodos para desarrollar y genotipar miles de marcadores en todo el genoma (Bao *et al.* 2020; Basak *et al.* 2019; Davey *et al.* 2011). Por ejemplo, la secuenciación de ADN asociada a sitios de restricción (ddRADseq), es una técnica que genera una gran cantidad de marcadores moleculares con polimorfismos de un solo nucleótido (SNPs) a partir de una sub-representación del genoma (Peterson *et al.* 2012; Li *et al.* 2019; Sudan *et al.* 2019; Bai *et al.* 2018; Davey *et al.* 2011). El método ddRAD-seq consiste en que el ADN genómico se digiere mediante la combinación de dos enzimas para luego secuenciar desde los sitios de restricción y así generar una gran cantidad de lecturas cortas, que luego son utilizadas para la búsqueda de los SNPs (Li *et al.* 2019; Sudan *et al.* 2019).

El método ddRAD-seq ha sido utilizado en estudios de diversidad genética en cultivos de tomate, cebolla, naranja y camote (Peterson *et al.* 2012; Lee *et al.* 2018; Shirasawa *et al.* 2016; Shirasawa *et al.* 2017; Penjor *et al.* 2016). A diferencia de la secuenciación del genoma completo, en ddRADseq solo se dirige a un subconjunto del genoma, pero con la ventaja de que no requiere ningún conocimiento previo del genoma de la especie de

interés, proporcionando buen rendimiento y rentabilidad para estudios de diversidad genética (Guillardín-Calvo *et al.* 2019; Pyne *et al.* 2017).

*Psidium friedrichsthalianum* (O. Berg) Nied es una especie cultivable de origen tropical de la familia Myrtaceae, subfamilia Myrtoideae, tribu Myrteae (Lucas *et al.* 2019). Produce un fruto conocido popularmente como “Cas”, “Guayaba ácida” o “Guayaba costarricense” (Barahona and Rivera 1995; Rojas-Rodríguez and Torres-Córdoba 2013) que es muy apetecido en la dieta de los costarricenses y centroamericanos (principalmente Nicaragua y El Salvador). En Costa Rica se vende en supermercados, “ferias del agricultor” y se utiliza principalmente para hacer refrescos pasteurizados, postres, mermeladas y otros productos relacionados (León 2000). Por lo general, esta especie se cultiva en grupos pequeños, jardines caseros, potreros, y más recientemente, en plantaciones de mediano tamaño, pero, sin apoyo técnico y promoción del cultivo. (Rojas-Rodríguez and Torres-Córdoba 2013). Por otra parte, esta especie cultivable se caracteriza por poseer flores hermafroditas, reproducción alógama y polinización principalmente realizada por insectos del orden Lepidóptera e Himenóptera (Barahona and Rivera 1995).

Algo destacable de esta especie cultivable es que su fruto ha sido potencializado como “alimento funcional”, debido a su alto contenido polifenólico (proantocianidinas (PACs) tipo A y B) y a su gran capacidad antioxidante (Rojas-Garbanzo *et al.* 2019; Cuadrado-Silva *et al.* 2017; Granados-Chinchilla *et al.* 2016; Flores *et al.* 2013; González *et al.* 2012). Un alimento funcional es aquel que además de sus componentes nutricionales tiene metabolitos secundarios que afectan positivamente la salud, más allá de los efectos nutricionales básicos, reduciendo el riesgo de enfermedades (Farag *et al.* 2020; Bigliardi & Galati 2013; Binns & Howlett 2009).

Investigaciones previas han demostrado el potencial del Cas como alimento funcional, por ejemplo, Flores *et al.* (2013), estudiaron en modelos *in vitro* de células HSAECs (Human Primary Small Airway Epithelial Cells) la capacidad que tienen los extractos del fruto de Cas para reducir la inflamación causada en pulmones por la exposición al humo de tabaco (EPOC). También, Granados *et al.* (2016) observaron actividad antimicrobiana contra bacterias comunes transmitidas por alimentos, utilizando aceites esenciales del fruto de Cas. Así mismo, González *et al.* (2012) compararon la capacidad antioxidante de 34 frutas tropicales, observando mayor capacidad antioxidante de los extractos del fruto de Cas en comparación con extractos de Guayaba blanca (white flesh), Dragon fruit, Noni, Mango, Naranja y Rambután. También, Rojas-Garbanzo *et al.* (2019) caracterizaron más de 100 compuestos del fruto de Cas, aislaron proantocianidinas (PACs) y flavonolignanos; y destacaron el potencial del fruto para comercializarlo como alimento funcional.

Actualmente, no hay información sobre la cantidad y distribución de la diversidad genética de poblaciones cultivadas de Cas en Costa Rica utilizando marcadores moleculares, la gran mayoría se han realizado en *Psidium guajava* y con marcadores tipo microsatélites (SSR); de igual manera, no se conoce los niveles de ploidía de estas poblaciones, Srivastava (1977) informó que este cultivo es diploide ( $2n = 2x = 11$ ) e Hirano (1967) identificó individuos tetraploides y hexaploides en muestras recolectadas en América Central; es probable que la diversidad en el número cromosómico previamente reportado para Cas puede ser una consecuencia de su proceso de domesticación en curso. Así mismo, tampoco se conoce el tamaño del genoma o valor C del Cas, se ha caracterizado solamente en congéneres, por ejemplo, *Psidium acutangulum*, *Psidium cattleianum*, *Psidium guajava* L. (cultivar blanco), *Psidium guajava* L. (cultivar rojo), *Psidium guineense* y *Psidium grandifolium* (Costa and Forni-

Martins 2006b, Costa *et al.* 2008; Machado-Marques *et al.* 2016; Coser *et al.* 2012; Souza *et al.* 2015). La ausencia de dicha información limita el desarrollo de programas de selección y mejoramiento genético en el cultivo de Cas, que sean eventualmente dirigidos a cosechar frutos de Cas con contenido alto y homogéneo de polifenoles, y que sea comercializado como alimento funcional.

Por lo tanto, el objetivo del presente estudio fue caracterizar diversidad genética de poblaciones cultivadas de Cas (*Psidium friedrichsthalianum* (O. Berg) Nied) en Costa Rica utilizando descriptores citogenéticos y moleculares para su eventual conservación y/o selección en programas de mejoramiento genético.

## Referencias

- Achigan-Dako EG, Fuchs J, Ahanchede A, Blattner FR. 2008. Flow cytometric analysis in *Lagenaria siceraria* (Cucurbitaceae) indicates correlation of genome size with usage types and growing elevation. *Plant Syst Evol.* 276:9–19.
- Awise JC. 1994. *Molecular Markers, Natural History and Evolution*. New York: Chapman and Hall.
- Bai B, Wang L, Zhang YJ. 2018. Developing genome-wide SNPs and constructing an ultrahigh-density linkage map in oil palm. *Sci Rep* 8, 691.
- Bao W, Ao D, Wuyun, T. 2020. Development of 85 SNP markers for the endangered plant species *Prunus mira* (Rosaceae) based on restriction site-associated DNA sequencing (RAD-seq). *Conservation Genet Resour.*
- Baraona M, Rivera G. 1995. Development of jocote (*Spondias purpurea* L.) and cas (*Psidium friedrichsthalianum* Ndz) in the premontano humid forest of Costa Rica. *Mesoamerican Agronomy.* 6: 23-31.
- Basak M, Uzun B, Yol E. 2019. Genetic diversity and population structure of the Mediterranean sesame core collection with use of genome-wide SNPs developed by double digest RAD-Seq. *PLoS ONE* 14(10): e0223757.
- Bigliardi B, Galati F. 2013. Innovation trends in the food industry: the case of functional foods. *Trends in Food Science & Technology.* 31:118–129.
- Binns N, Howlett J. 2009. Functional foods in Europe: international developments in science and health claims. *European Journal of Nutrition.* 48: S3–S13.
- Birchler JA. 2013. Genetic rules of heterosis in plants. In: *polyploid and hybrid genomics*. Wiley. pp 313-321.

- Bordoni P, Hodgkin T. 2012. Climate change and the conservation of plant genetic resources. *J Crop Improv.* 26(3):329–45.
- Bretting, P. K., and M. P. Widrechner. 1995. Genetic markers and plant genetic resource management. *Plant Breed. Rev.* 31:11–86.
- Chaveerach A, Sudmoon R, Tanee T, Mokkamul P and Tanomtong A. 2007. Genetic relationships in a population of *Nelumbo nucifera* Gaertn (Nelumbonaceae). *Journal of Biological Sciences* 7: 1388–1393.
- Coser SM, Ferreira MFS, Ferreira A, Mitre LK, Carvalho CR, Clarindo WR. 2012. Assessment of genetic diversity in *Psidium guajava* L. using different approaches. *Scientia Horticulturae.* 148: 223–229.
- Costa IR, Dornelas MC, Forni-Martins ER. 2008. Nuclear genome size variation in fleshy-fruited Neotropical Myrtaceae. *Plant Systematics and Evolution.* 276: 209-217.
- Costa IR, Forni-Martins ER. 2006b. Chromosome studies in species of *Eugenia*, *Myrciaria* and *Plinia* (Myrtaceae) from south-eastern Brazil. *Australian Journal of Botany.* 54: 409-415.
- Cuadrado-Silva CT, Pozo-Bayón MÁ, Osorio C. 2017. Targeted Metabolomic Analysis of Polyphenols with Antioxidant Activity in Sour Guava (*Psidium friedrichsthalianum* Ndz.). *Fruit Molecules.* 22 (1): 11.
- Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. 2011. Special features of RAD sequencing data: Implications for genotyping. *Molecular Ecology.* 22: 3151–3164.
- Doležel J, Bartoš J. 2005. Flow cytometry of plant DNA and estimation of nuclear genome size. *Ann Bot.* 95 (1): 99–110.
- El-Esawi MA. 2012. Assessing the genetic diversity, phylogenetic relationships, and disease resistance genes in Irish Brassica oleracea species. PhD Thesis, Dublin Institute of Technology, Ireland.

- Farag MA, Abdelwareth A, Sallam IE, el Shorbagi M, Jehmlich N, Fritz-Wallace K, Schäpe S, Rolle-Kampczyk U, Ehrlich A, Wess LA, von Bergen M. 2020. Metabolomics reveals impact of seven functional foods on metabolic pathways in a gut microbiota model. *Journal of Advanced Research*. 23: 47–59.
- Flores G, Dastmalchi K, Wu SB, Whalen K, Dabo AJ, Reynertson K. 2013. Phenol-rich extract of the Costa Rican guava pulp (*Psidium friedrichsthalianum* Ndz) with antioxidant and anti-inflammatory activity. Potential for COPD therapy. *Food Chemistry*. 14: 889-895.
- González E, Vaillant F, Pérez AG, Rojas T. 2012. Antioxidant protection mediated by in vitro cells of human erythrocytes by some common tropical fruits. *J Nutr Food Sci*. 2: 139.
- Granados-Chinchilla F, Villegas E, Molina A, Arias C. 2016. Composition, Chemical Fingerprinting and Antimicrobial Assessment of Costa Rican Cultivated Guavas (*Psidium friedrichsthalianum* (O. Berg) Ndz and *Psidium guajava* L.) Essential Oils from Leaves and Fruits. *Nat Prod Chem Res*. 4:5.
- Greilhuber JJ, Dolezel MA, Lysak MD. 2005. The origin, evolution and proposed stabilization of the terms "genome size" and "C value" to describe the contents of nuclear DNA. *Ann. Bot*. 95: 255-260.
- Guillardín-Calvo L, Mora-Márquez F, Soto Á. 2019. RADDESIGNER: a workflow to select the optimal sequencing methodology in genotyping experiments on woody plant species. *Tree Genetics & Genomes*. 15: 64.
- Hirano RT. 1967. Chromosomal and pollination studies as related to intra-specific and interspecific compatibility in the genus *Psidium*. Master's thesis, University of Hawaii, Honolulu.
- Jasienski M, Bazzaz FA. 1995. Genome size and high CO<sub>2</sub>. *Nature*. 376:559–560.



- Knight CA, Ackerly DD. 2002. Variation in nuclear DNA content across environmental gradients: A quantile regression analysis. *Ecol Lett.* 5:66–76.
- Kumar J, Agrawal V. 2017. Analysis of genetic diversity and population genetic structure in *Simarouba glauca* DC. (an important bio-energy crop) employing ISSR and SRAP markers Jatin Kumar, Veena Agrawal. *Industrial Crops and Products* 100: 198–207.
- Kumar S, Banks TW, Cloutier S. 2017. SNP discovery through next-generation sequencing and its applications. *Int J Plant Genomics*.34: 45-58.
- Laguado, J. 2007. Applications of flow cytometry in veterinary microbiology and agriculture. *Rev MVZ Córdoba* 12(2):1077–1095.
- Lee JH, Natarajan S, Biswas MK, Shirasawa K, Isobe S, Kim HT. 2018. SNP discovery of Korean short day onion inbred lines using double digest restriction site-associated DNA sequencing. *PLoS ONE* 13(8): e0201229
- Leitch AR, Leitch IJ. 2008. Genomic Plasticity and the Diversity of Polyploid Plants. *Science* 320(5875):481–483.
- León J.2000. Botany of tropical crops. San José, Costa Rica: Ed. Ilca. pp. 445. Mathers JC. 2006. Plant foods for human health: research challenges. *Proc Nutr Soc* 65: 198–203.
- Li S, Lv S, Yu K. 2019. Construction of a high-density genetic map of tree peony (*Paeonia suffruticosa* Andr. Moutan) using restriction site associated DNA sequencing (RADseq) approach. *Tree Genetics & Genomes* 15, 63.
- Lin BZ, Sasazaki S, Mannen H. 2010. Genetic diversity and structure in *Bos taurus* and *Bos indicus* populations analyzed by SNP markers. *Anim. Sci. J.* 81 281–289.

- Lucas EJ, Holst B, Sobral M, Mazine FF, Nic Lughadha EM, Barnes Proença CE, Ribeiro da Costa I, Vasconcelos TNC. 2019. A New Subtribal Classification of Tribe Myrteae (Myrtaceae). *Systematic Botany*. 44(3): pp. 560–569.
- Machado-Marques A, Tuler AC, Carvalho CR, Tavares-Carrijo T, Silva-Ferreira MF, Clarindo WR. 2016. Refinement of the karyological aspects of *Psidium guineense* (Swartz, 1788): a comparison with *Psidium guajava* (Linnaeus, 1753). *CompCytogen* 10 (1): 117-128.
- Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S. 2012. SNP markers and their impact on plant breeding. *Int J Plant Genomics*. (30): 23-36.
- McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W. 2008. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. *BMC Genetics* 9:37.
- Meagher TR, Costich DE. 1994. Sexual dimorphism in nuclear DNA content and floral morphology in populations of *Silene latifolia* (Caryophyllaceae). *Am J Bot*. 81:1198–1204.
- Olsen KM, Wendel JF. 2013. A Bountiful Harvest: Genomic Insights into Crop Domestication Phenotypes. *Annu Rev Plant. Biol.* 64: 47–70.
- Penjor T, Mimura T, Kotoda N, Matsumoto R, Nagano JA, Honjo NM. 2016. RAD-Seq analysis of typical and minor Citrus accessions, including Bhutanese varieties. *Breed Sci.* 66(5):797–807.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7: 37-45.
- Pyne R, Honig J, Vaiciunas J, Koroch A, Wyenandt C, Bonos S, Simon J. 2017. A first linkage map and downy mildew resistance QTL discovery for sweet basil (*Ocimum basilicum*) facilitated by double digestion restriction site associated DNA sequencing (ddRADseq). *PloS one*. 12(9).

- Rojas-Garbanzo C, Wintera J, Montero ML, Zimmermann BF, Schieber A. 2019. Characterization of phytochemicals in Costa Rican guava (*Psidium friedrichsthalianum* Ndz) Fruit and stability of main compounds during juice processing - (U) HPLC-DAD- ESI-TQD-MS. *Journal of Food Composition and Analysis*. 75: 26-42.
- Rojas-Rodríguez F, Torres-Córdoba G. 2013. Trees of the Central Valley of Costa Rica: reproduction Cas (*Psidium friedrichsthalianum* (Berg) Ndz). *Kurú Mesoamerican Forest Magazine* (Costa Rica). 120 (10): 1320-1323.
- Shirasawa K, Hirakawa H, Isobe S. 2016. Analytical work flow of double-digest restriction site-associated DNA sequencing based on empirical and in silico optimization in tomato. 23(2):145–153.
- Shirasawa K, Tanaka M, Takahata Y, Ma D, Cao Q, Liu Q. 2017. A high-density SNP genetic map consisting of a complete set of homologous groups in autohexaploid sweetpotato (*Ipomoea batatas*). *SciRep*. 7:44207.
- Smarda P, Bures P, Horova L, Rotreklova O. 2008. Intrapopulation genome size dynamics in *Festuca pallens*. *Ann Bot*. 102:599–607.
- Souza AG, Resende LV, Lima IP, Martins LSS, Techio VH. 2015. Chromosome number and nuclear DNA amount in *Psidium* spp. resistant and susceptible to *Meloidogyne enterolobii* and its relation with compatibility between rootstocks and commercial varieties of guava tree. *Plant Systematics and Evolution*. 301: 231-237.
- Srivastava HC. 1977. Cytological studies in *Psidium friedrichsthalianum* Ndz. *Cytologia*. 42: 395-400.
- Sudan J, Singh R, Sharma S. 2019. ddRAD sequencing-based identification of inter-genepool SNPs and association analysis in *Brassica juncea*. *BMC Plant Biol* 19, 594.

- Thavamanikumar S, McManus LJ, Tibbits JFG, Bossinger G. 2011. The significance of single nucleotide polymorphisms (SNPs) in *Eucalyptus globulus* breeding programs. *Australian Forestry*. 74:23-29.
- Treuren R, Hintum TJL. 2014. Next-generation genebanking: plant genetic resources management and utilization in the sequencing era. *Plant Genet Resour*. 12:298–307.
- Vinogradov AE. 1995. Nucleotypic effect in homeotherms: Body-mass-corrected basal metabolic rate of mammals is related to genome size. *Evolution*. 49:1249–1259.
- Wakamiya I. 1993. Genome size and environmental factors in the genus *Pinus*. *Am J Bot*. 80:1235–1241.
- Washburn JD, Birchler JA. 2014. Polyploids as a “model system” for the study of heterosis. *Plant Reprod*. 27:1–5.
- Wendel JF, Jackson SA, Meyers BC, Wing RA. 2016. Evolution of plant genome architecture. *Genome Biol*. 17:37.

**CAPÍTULO 1.** Genome size and chromosome number of *Psidium*  
*friedrichsthalianum* (O. Berg) Nied (“Cas”) in six populations of Costa Rica  
(con formato para International Journal of Cytology, Cytosystematics and Cytogenetics)

**Genome size and chromosome number of *Psidium friedrichsthalianum* (O. Berg) Nied (“Cas”) in six populations of Costa Rica**

Mónica Rojas-Gómez<sup>a\*</sup>, Alfonso García-Piñeres<sup>b</sup>, Pablo Bolaños-Villegas<sup>c,d</sup>, Griselda Arrieta-Espinoza<sup>e</sup> and Eric J. Fuchs<sup>f</sup>

<sup>a</sup> *Centro Nacional de Innovaciones Biotecnológicas (CENIBiot), CeNAT-CONARE, 1174-1200 San José, Costa Rica.*

<sup>b</sup> *Centro de Investigación en Biología Celular y Molecular and Escuela de Química, Universidad de Costa Rica, 11501-2060 San José (Costa Rica).*

<sup>c</sup> *Fabio Baudrit Agricultural Research Station, Universidad de Costa Rica, La Garita, Alajuela 20101, Costa Rica.*

<sup>d</sup> *Lankester Botanical Garden, University of Costa Rica, P.O. Box 302-7050, Cartago, Costa Rica*

<sup>e</sup> *Centro de Investigación en Biología Celular y Molecular (CIBCM), Universidad de Costa Rica 11501-2060 San José, Costa Rica.*

<sup>f</sup> *Escuela de Biología, Universidad de Costa Rica, 11501-2060 San José, Costa Rica.*

Corresponding author: Mónica Rojas-Gómez, ORCID: 0000-0002-0245-2911, Centro Nacional de Innovaciones Biotecnológicas (CENIBiot), CeNAT-CONARE, 1174-1200 San José, Costa Rica, emails: [morogo27@gmail.com](mailto:morogo27@gmail.com); [mrojas@cenat.ac.cr](mailto:mrojas@cenat.ac.cr).

Alfonso García- Piñeres: ORCID: 0000-0002-3000-4630

Pablo Bolaños-Villegas: ORCID: 0000-0003-1729-0561

Griselda Arrieta-Espinoza: ORCID: 0000-0001-9235-5834

Eric J. Fuchs Castillo: ORCID: 0000-0002-6645-9602

# **Genome size and chromosome number of *Psidium friedrichsthalianum* (O. Berg) Nied (“Cas”) in six populations of Costa Rica**

## **Abstract**

*Psidium friedrichsthalianum* (O. Berg) Nied is a species found from southern Mexico, Central America; and there are reports that it is also found in Venezuela and Ecuador. It is a common fruit component of the Costa Rican diet, and it is valued industrially for its high content of polyphenols, mainly proanthocyanidins (PACs). This crop is not completely domesticated and there are no improved varieties produced through plant breeding. Genome size or ploidy levels have not been investigated in Costa Rican populations of *Psidium friedrichsthalianum*. Information about chromosome number and genome size is paramount for plant breeding strategies. Therefore, the main objective of our study was to determine chromosome number using pollen meiocytes and genome size by flow cytometry in six populations of *P. friedrichsthalianum* in Costa Rica. We found  $x = 11$  bivalent chromosomes in all meiocytes analysed, classifying these populations as diploid. All populations had an average nuclear DNA content of  $2C = 1.960 \pm 0.005$  pg. No statistically significant differences in nuclear DNA content were found among populations. We conclude that the consistency in chromosome number and genome size among populations suggests a common origin among them. Our estimates of the number of chromosomes and genome size of *P. friedrichsthalianum* determined in this study will be essential for future breeding programs, hybridization practices and development of QTL (Quantitative Trait Loci).

Keywords: ploidy; fluorescent microscopy; 2C nuclear DNA; flow cytometry; Costa Rican Guava; plant breeding.

Word Count: 2546

## Introduction

*Psidium friedrichsthalianum* (O. Berg) Nied is a tropical species in the family Myrtaceae, subfamily Myrtoideae, tribe Myrteae (Lucas *et al.* 2019); commonly known as "Cas", "Sour Guava" or "Costa Rican Guava". It is a medium-sized tree with reddish branches and abundant foliage of intense green color. Flowers are perfect, possibly allogamous and pollination is performed by bees and occasionally by hummingbirds (Barahona and Rivera 1995). Fruits are fleshy globose berries, between 5 and 10 cm in diameter with a greenish to yellow exocarp and a very distinct soft and acidic pulp. In addition, it is presumed that its center of origin is in Costa Rica (Barahona and Rivera 1995; Rojas-Rodríguez and Torres-Córdoba 2013). "Cas" fruits are characterised by abundant polyphenol content, mainly proanthocyanidins (PACs); these metabolites have important antioxidant, anti-inflammatory, antimicrobial and vasodilatory properties (Cuadrado-Silva *et al.* 2017; Flores *et al.* 2013; Rojas-Garbanzo *et al.* 2019; Granados-Chinchilla *et al.* 2016; González *et al.* 2012). Vasconcelos *et al.* (2019) described the chemical composition and allelopathic properties of essential oils extracted from *P. friedrichsthalianum*, suggesting that this oil may be used as a natural weed control comparable in efficacy, to synthetic herbicides. This fruit is considered an important resource due to its photochemical properties; however, few studies have been conducted on this tropical fruit.

The germplasm of *P. friedrichsthalianum* in Costa Rica has not yet been genetically characterized, however, Srivastava (1977), reported this species as diploid ( $2n = 2x = 11$ ), while Hirano (1967) reported tetraploid and hexaploid individuals in Central America samples. The diversity in chromosomal number previously reported for *P. friedrichsthalianum* may be a consequence of its ongoing domestication process. Information on chromosome numbers in the Myrtaceae is generally scarce, the fairly



small chromosomes found in this taxonomic group, which rarely exceed 2 mm (Costa 2004), may curb chromosome determination. Presently, genome sizes have been reported for *Psidium acutangulum*, *Psidium cattleianum*, *Psidium guajava* L. (white cultivar), *Psidium guajava* L. (red cultivar), *Psidium guineense* and *Psidium grandifolium* (Costa and Forni-Martins 2006b, Costa *et al.* 2008; Machado-Marques *et al.* 2016; Coser *et al.* 2012; Souza *et al.* 2015) (Table 1). However, the genome size or the 2C value of *P. friedrichsthalianum* have not been analyzed yet.

Estimates of the number of chromosomes and genome size for *P. friedrichsthalianum* are essential for the design of effective improvement strategies, such as hybridization practices, the development of QTL (Quantitative Trait Loci), as well as to better understand the effects of inbreeding and heterosis (Birchler 2013; Washburn and Birchler 2014). Additionally, with the current development of second and third generation sequencing techniques (NGS), information on genome size or C values are essential to establish appropriate experimental conditions, to effectively prepare genomic libraries and sequencing of complete genomes (Leitch and Leitch 2008). The C-values reported here may be used as a tool for genomic analysis in this species, which should benefit genetic improvement practices in this species.

Therefore, given the absence of information on ploidy level and nuclear DNA content in populations of *Psidium friedrichsthalianum*; we aimed to determine the chromosome number via fluorescent DAPI stain and flow cytometry to determine the nuclear DNA content of this tropical fruit in six populations of Costa Rica, its likely centre of origin.

## **Materials and Methods**

### ***Sample Collection***

We analysed individuals from six populations of *P. friedrichsthalianum* in Costa Rica. Samples were collected from local small-scale plantations from different regions in the country (Table 2). Plantations were located at different elevations ranging from sea level to over 1500 m asl (metres above sea level). Samples were always taken from reproductive trees and care was taken to collect samples from individuals that were separated by at least 10 meters to avoid collecting possible genets.

### ***Chromosomal count using DAPI stain***

Chromosome counts were performed on pollen mother cells in meiotic metaphase. At least seven flower buds were collected from each of six populations; flower buds ranged between 0.7 cm and 0.8 cm in length. Flower buds were fixed in FAA solution (96% ethanol, 5% glacial acetic acid and 40% formaldehyde) for 24 hours. As suggested by Dyer (1979), flower buds were dissected to 3/4 of their final bud size. Anthers were placed on slides and subjected to mechanical disaggregation (macerating anthers with a thin spatula) adding occasional drops of acetic acid to prevent desiccation. Macerated anthers were stained with DAPI fluorochrome (Sigma-Aldrich, Illinois, USA) and were incubated in the dark for 5 min. We used an epifluorescence microscope (Olympus BX50, Olympus Corporation, Tokyo, Japan) to visualize and photograph stained cells. Only cells that were in a state of meiotic metaphase were photographed. For each population at least five slides were evaluated and at least one cell was recorded in a state of meiotic metaphase. Finally, to facilitate chromosome counts, image color adjustments were performed with Adobe Photoshop CS82 (Adobe Systems, San Jose, CA) and the ImageJ software (US National Institutes of Health, 2007) was used to count chromosomes.

### ***Flow cytometry estimates of DNA content***

We used flow cytometry to estimate genome size on *P. friedrichsthalianum*. We collected fruits from at least 10 individuals per population. Seeds were manually extracted, washed with tap water and dried in open air for a week to remove moisture. After one week, seeds were sown in pots with vermiculite soil and placed in a greenhouse for 12 weeks until seedling emergence. Ten seedlings per population were analyzed in a BD FACSCalibur TM (Becton Dickinson, San Jose, CA, USA) flow cytometer. After initial parameter adjustment in the flow cytometry equipment, samples were prepared following the protocol by Dolezel *et al.* (2007) with some modifications. We used 1 mg of leaf sample from *Glycine max* as a reference ( $2C = 2.50$  pg) (Dolezel *et al.* 2007) and 5 mg of leaf tissue from *P. friedrichsthalianum*. Leaves were placed in a petri dish on ice, then 1ml of OTTO-I lysis buffer (0.1 M citric acid, 0.5% (vol / vol) Tween 20) supplemented with 2 mM dithiothreitol (DTT) was added to the leaf cutouts (Otto 1990). Subsequently, leaves were cut with a razor blade until homogenization. The extract was filtered through a 41µm Nylon mesh, onto a 2.0 ml microcentrifuge tube. The filtrate was centrifuged at 10,000 rpm for 5 minutes. The supernatant was discarded and the pellet was resuspended in 100 µl of OTTO-I lysis buffer and incubated for 15 min at 4 ° C. After incubation and prior to analysis in a flow cytometer, 300 µl of OTTO-II buffer (0.4 M Na<sub>2</sub>HPO<sub>4</sub> · 12 H<sub>2</sub>O) (Otto 1990), 20 µl of propidium iodide (50 µg / ml) and 2 µl of RNase (50 µg / ml) were added to the mixture.

All measurements were based on the fluorescence of at least 5000 total events (total nuclei). We analyzed two independent replicas of each sample on different days and estimated an average nuclear DNA content. Mean fluorescence intensity (MFI), number of events per peak and variation coefficient were all calculated using the FCS

Express 4 Flow Cytometry software (De Novo Software, Los Angeles, CA). Finally, the nuclear DNA content was calculated according to Dolezel *et al.* (2007) as follows:

$$A = \frac{(B \times C)}{D}$$

Where A = 2C (pg) nuclear DNA content concentration of *P. friedrichsthalianum*; B = Mean fluorescence intensity (MFI) of the G<sub>0</sub> / G<sub>1</sub> peak of *P. friedrichsthalianum*; C = 2C (pg) nuclear DNA content of the internal standard; D = MFI of the G<sub>0</sub> / G<sub>1</sub> peak of the internal standard. Genome size was estimated from DNA content as 1 picogram (pg) of DNA being equivalent to 978 megabase pairs (Mbp) (Bennett *et al.* 2000; Dolezel *et al.* 2003). The 2C nuclear DNA content data of all individuals was compared among populations with a one-way ANOVA, followed by Tukey's test to determine individual differences (p < 0.05). Statistics were done using R 3.5.0 software (R Core Team 2018).

## Results and Discussion

We consistently found 11 bivalent chromosomes in all meiocytes from *Psidium friedrichsthalianum* (Fig 2) across all populations (Table 2). Taking into account that the basic chromosome number of the Myrtaceae family is x = 11 (Atchison 1947; Raven 1975) we classified all Costa Rican samples of *P. friedrichsthalianum* as diploid (2n=2x=22). The diploid nature of the Costa Rican guava mirrors the results from Srivastava (1977), who similarly found a 2n = 2x = 22 diploid chromosome count in different genotypes of *Psidium friedrichsthalianum*. Costa and Forni-Martins (2006a, b, 2007) also described chromosome numbers for 50 species in the Myrtaceae, and found a predominance of 2n = 2x = 22 diploid species. Naitani and Srivastava (1965), Coser *et al.* (2012), Éder-Silva *et al.* (2007), and Souza *et al.* (2015), all found predominantly diploid species in the *Psidium* genus, such as in *Psidium chinense* and *Psidium guajava*.

Previous results reported *P. friedrichsthalianum* individuals with  $2n=4x=44$  and even  $2n=6x=66$  (Hirano 1967), suggesting that this species may have tetraploid and hexaploid members. These results clearly indicate that there may be variation in ploidy levels among populations of *P. freidrichsthalianum* in different areas. In contrast, our results show that at least in Costa Rica, cultivated populations are consistently diploid. This chromosomal uniformity may be the result of a common historical origin among populations. Alternatively, our results may also be a consequence of artificial selection by farmers who selected cytotypes with specific homogenous traits of interests such as fruit size of pulp content. Multiple cytotypes have also been found in other *Psidium* congeners, for example in *Psidium cattleianum* the cytotypes  $2n = 44, 66, 77$  and  $88$  have been described (Costa and Forni-Martins 2006a; Costa 2009). Multiple cytotypes have been also found in populations of *Psidium guineense* and *Psidium guajava* (Srivastava 1977; Costa and Forni-Martins 2006a; Éder-Silva *et al.* 2007; Souza *et al.* 2015). Polyploidy is recognized as one of the main evolutionary forces in angiosperms (Soltis *et al.* 2015); and it is frequently associated with interspecific hybridization followed by chromosomal duplication to restore hybrid fertility (Soltis *et al.* 2009). Results from congeners suggests that *P. friedrichsthalianum* may also have the potential to create other cytotypes may represent important prospects for future breeding programs.

Our study found bivalent and univalent chromosomes in meiocytes of *P. freidrichsthalianum* (Fig 2 a, b, c). Chromosomes were also observed in a trivalent state (Fig 2d) and this is consistent with previous observations by Srivastava (1977) in this species. Univalent chromosomes are frequently observed in plants; these can arise through three different ways: (i) when a chromosome is not matched completely in zygotene stage; (ii) when paired bivalents separate in diplotene because robust chiasmata have not yet formed between them; (iii) due to premature disjunction of the bivalents

during anaphase (Pires-Bione *et al.* 2000). The premature migration of univalent chromosomes to the poles during cell division is common in plants, giving rise to micronuclei (Pagliarini 1990; Pagliarini and Pereira 1992; Consolaro *et al.* 1996). Alternatively, univalent chromosomes may occasionally occur in plants due to environmental factors such as temperature fluctuations (Heilborn 1934; Katayama 1935). Some of our sites differ drastically in climatic conditions, however, further studies are needed to better understand the cytology of this species.

Our flow cytometry estimates were very consistent across all plant samples. Our coefficients of variation were all less than 5% (Table 3), which confirms that our suspensions had a sufficient number of stoichiometrically stained and intact nuclei. Additionally, DTT used in nuclei suspensions proved to be effective inhibiting cytosolic interfering compounds which resulted in clear histograms. DTT is commonly used in flow cytometry studies because of its broad antioxidant activity (Dolezel *et al.* 2007). In our study, DTT was very efficient because many woody species in the Myrtaceae, as is the case of *P. friedrichsthalianum*, contain abundant secondary metabolites that may interfere with DNA content staining (Loureiro *et al.* 2006) (Ohri and Kumar 1986).

We determined a mean nuclear value of  $2C = 1.960 \pm 0.005$  pg for *P. friedrichsthalianum*, equivalent to 1916.88 Mbp (Bennett *et al.* 2000) (Fig 3, Table 3). Nuclear DNA content did not statistically vary among all six populations ( $F=0.29$ ;  $df=5$ ;  $p = 0.917$ ). Leitch *et al.* (1998) and Soltis *et al.* (2003) classified species with  $1C \leq 1.4$  pg content as species with a very small genomes compared to other angiosperms. Therefore, given our 1C estimates ( $1C = 0.98 \pm 0.005$  pg) (Table 3) the Costa Rican guava should also be classified as a small genome species. Consistently, Machado-Marques *et al.* (2016) found that *Psidium guajava* and *Psidium guineense*, also have very small genomes as  $1C = 0.475$  pg and  $1C = 0.925$  pg, respectively (Table 1). Almeida- Carvalho

*et al.* (2017) determined that 25 species of the genus *Eucalyptus* (Myrtaceae), all had 1C values between 1C=0.40 pg and 1C=0.75 pg which may indicate that the Myrtaceae family may typically contain species with smaller genomes. On the other hand, our 2C estimates ( $2C = 1.960 \pm 0.005$  pg) are within the range described by Souza *et al.* (2015), who also used flow cytometry on different species of *Psidium* and found 2C values that ranged between 2C=0.99 pg and 2C=5.48 pg. However our estimates are significantly higher than those found for different varieties of *Psidium guajava*; for example, Coser *et al.* (2012) found  $2C = 0.507$  pg for the white varieties, and  $2C = 0.551$  pg or  $2C = 0.950$  pg for the red varieties; while Souza *et al.* (2015) found  $2C = 0.990$  pg and  $2C = 1.020$  pg in purple and "Paluma" varieties respectively (Table 1). These differences in genome size may be due to (i) natural or bred adaptations of these species to different environmental conditions (Cavallini and Natali 1990), for example, to new cultivation environments; (ii) hybridization events, or (iii) changes in repetitive DNA sequences (Martel *et al.* 1997). Several authors have suggested that transposable elements (TE) may be important in the evolution of genome sizes in plants (Wang *et al.* 2016; Wendel *et al.* 2016; Zhao *et al.* 2016). For example, Almeida-Carvalho *et al.* (2017) compared the genome size of two *Eucalyptus* species, *E. botryoides* ( $2C=1.350$  pg) and *E. microcorys* ( $2C=1.040$  pg); and found big differences in genome size between them, although both had the same chromosome number ( $2n = 2x = 22$ ) (Table 1). They argued that variations in 2C values in these *Eucalyptus* were caused by chromosome rearrangement and possibly TE elements.

Therefore, our results on ploidy level and genome size of *P. friedrichsthalianum*, contribute to the cytogenetic characterization of this economically important fruit species. This information may be used to design regional conservation strategies that preserve local genetic resources. Flow cytometry may be used to assess ploidy level in *in vitro*

propagated plants (Ochatt *et al.* 2011), to screen for plants with higher ploidy levels, which may have new features of economic interest such as increased fruit size, or better juicing capabilities. Additionally, results from our study could aid the taxonomic definition of *P. friedrichsthalianum* species and the understanding of phylogenetic relationships among other members in the genus *Psidium*.

## Conclusions

Populations of *Psidium friedrichsthalianum* from six different regions of Costa Rica, the likely centre of origin of this species, have a chromosome number equal to  $2n = 2x = 22$ , indicating that cultivated populations in Costa Rica, are all consistently diploid. Furthermore, these populations have an average 2C nuclear DNA content of  $1.960 \pm 0.005$  pg. The uniformity found across populations in terms of chromosomal number and nuclear DNA content, suggests a common origin among them.

## Acknowledgements

We acknowledge the help of Estación Experimental Agrícola Fabio Baudrit Moreno (Universidad de Costa Rica); Centro Nacional de Innovaciones Biotecnológicas (CENIBiot) at CeNAT-CONARE; Finca Los Diamantes at Instituto Nacional de Innovación y Transferencia en Tecnología Agropecuaria (INTA). This work would not have been possible without local “Cas” farmers who provided free samples from their farms: David Badilla, Gregorio Menocal, Juan Pablo Orozco, Alfonso Ruíz, José Fuentes. This work was conducted under permit # 133-2018 from the Comisión de Biodiversidad-UCR at Universidad de Costa Rica.



## Disclosure Statement

The authors declare that we do not have any conflict or competing interests with funding agencies.

## Geolocation data

Geolocation data is found on Table 2.

## Funding details

This work was supported by the Vicerrectoría de Investigación de la Universidad de Costa Rica (UCR) under grant 111-B7-261; Consejo Nacional de Rectores (CONARE) under grant FEES-15-2019.

## References

- Almeida-Carvalho GM, Carvalho CR, Ferrari-Soares FA. 2017. Flow cytometry and cytogenetic tools in eucalypts: genome size variation  $\times$  karyotype stability. *Tree Genetics & Genomes*. 13: 106
- Atchison E. 1947. Chromosome numbers in the Myrtaceae. *American Journal of Botany*. 34: 159-164.
- Barahona M, Rivera G. 1995. Development of jocote (*Spondias purpurea* L.) and cas (*Psidium friedrichsthalianum* Niedz) in the premontano humid forest of Costa Rica. *Mesoamerican Agronomy*. 6: 23-31.
- Bennett MD, Bhandol P, Leitch IJ. 2000. Nuclear DNA amounts in angiosperms and their modern uses - 807 new estimates. *Annals of Botany*. 86: 859-909.
- Birchler JA. 2013. Genetic rules of heterosis in plants. In: *polyploid and hybrid genomics*. Wiley. pp 313-321
- Cavallini A, Natali L. 1990. Nuclear DNA variability within *Pisum sativum* (Leguminosae):

- cytometric analyzes. *Plant Systematics and Evolution*. 173: 179–185.
- Consolaro MEL, Pagliarini MS, Chaves LJ. 1996. Meiotic behavior, pollen fertility and seed production in Brazilian populations of *Centella asiatica* (L.) Urban (Umbelliferae). *Cytologia*. 61: 375–381.
- Coser SM, Ferreira MFS, Ferreira A, Mitre LK, Carvalho CR, Clarindo WR. 2012. Assessment of genetic diversity in *Psidium guajava* L. using different approaches. *Scientia Horticulturae*. 148: 223–229.
- Costa IR, Dornelas MC, Forni-Martins ER. 2008. Nuclear genome size variation in fleshy-fruited Neotropical Myrtaceae. *Plant Systematics and Evolution*. 276: 209–217.
- Costa IR, Forni-Martins ER. 2006a. Chromosome studies in Brazilian species of *Campomanesia* Ruiz & Pavon and *Psidium* L. (Myrtaceae Juss.). *Caryologia*. 59: 7–13.
- Costa IR, Forni-Martins ER. 2006b. Chromosome studies in species of *Eugenia*, *Myrciaria* and *Plinia* (Myrtaceae) from south-eastern Brazil. *Australian Journal of Botany*. 54: 409–415.
- Costa IR, Forni-Martins ER. 2007. Karyotype analysis in South American species of Myrtaceae. *Botanical Journal of the Linnean Society*. 155 (4): 571–580.
- Costa IR. 2004. Estudos cromossômicos em espécies de Myrtaceae Juss. no sudeste do Brasil. Master Thesis. Universidade Estadual de Campinas (UNICAMP). Campinas, SP.
- Cuadrado-Silva CT, Pozo-Bayón MÁ, Osorio C. 2017. Targeted Metabolomic Analysis of Polyphenols with Antioxidant Activity in Sour Guava (*Psidium friedrichsthalianum* Ndz.). *Fruit Molecules*. 22 (1): 11.
- Dolezel J, Bartos J, Voglmayr H, Greilhuber J. 2003. Nuclear DNA content and genome size of trout and human. *Cytometry*. 51:127–128
- Dolezel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nature protocols*. 2 (9): 2233–2244.

- Dyer AF. 1979. The culture of fern gametophytes for experimental investigation. In A.F. Dyer(ed.) The experimental biology of ferns. Academic Press, London. pp.253-305.
- Éder-Silva E, Felix LP, Bruno RLA. 2007. Cytogenetics of some fruit species native to northeastern Brazil. Brasileira Fruticultura Magazine Jaboticabal. 29: 110-114.
- Flores G, Dastmalchi K, Wu SB, Whalen K, Dabo AJ, Reynertson K. 2013. Phenol-rich extract of the Costa Rican guava pulp (*Psidium friedrichsthalianum* Ndz) with antioxidant and anti-inflammatory activity. Potential for COPD therapy. Food Chemistry. 14: 889-895.
- González E, Vaillant F, Pérez AG, Rojas T. 2012. Antioxidant protection mediated by in vitro cells of human erythrocytes by some common tropical fruits. J Nutr Food Sci. 2: 139.
- Granados-Chinchilla F, Villegas E, Molina A, Arias C. 2016. Composition, Chemical Fingerprinting and Antimicrobial Assessment of Costa Rican Cultivated Guavas (*Psidium friedrichsthalianum* (O. Berg) Ndz and *Psidium guajava* L.) Essential Oils from Leaves and Fruits. Nat Prod Chem Res. 4:5.
- Greilhuber J, Obermayer R. 1997. Genome size and maturity group in *Glycine max* (soybean). Heredity. 78: 547-551.
- Heilborn O. 1934. On the origin and preservation of polyploidy. Hereditas. 19: 233-242.
- Hirano RT. 1967. Chromosomal and pollination studies as related to intra-specific and interspecific compatibility in the genus *Psidium*. Master's thesis, University of Hawaii, Honolulu.
- Katayama Y. 1935. Further investigations on synthesized octoploid *Aegilotrichum*. Jour. Coll. Agr. Imp. Univ. Tokyo 13: 397-414.
- Leitch IJ, Chase MW, Bennett MD. 1998. Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. Ann Bot 82 (suppl. A): 85–94.

- Leitch AR, Leitch IJ. 2008. Genomic Plasticity and the Diversity of Polyploid Plants. *Science* 320(5875):481–483.
- León J. 2000. Botany of tropical crops. San José, Costa Rica: Ed. Ilca. pp. 445.
- Loureiro J, Rodríguez E, Dolezel J, Santos C. 2006. Comparison of four nuclear isolation buffers for plant DNA flow cytometry. *Ann Bot London*. 98: 679-689.
- Lucas EJ, Holst B, Sobral M, Mazine FF, Nic Lughadha EM, Barnes Proença CE, Ribeiro da Costa I, Vasconcelos TNC. 2019. A New Subtribal Classification of Tribe Myrteae (Myrtaceae). *Systematic Botany*. 44(3): pp. 560–569.
- Machado-Marques A, Tuler AC, Carvalho CR, Tavares-Carrijo T, Silva-Ferreira MF, Clarindo WR. 2016. Refinement of the karyological aspects of *Psidium guineense* (Swartz, 1788): a comparison with *Psidium guajava* (Linnaeus, 1753). *CompCytogen* 10 (1): 117-128.
- Martel E, De Nay D, Siljak-Yakoviev S, Brown S, Sarr A. 1997. Genome size variation and basic chromosome number in pearl millet and fourteen related *Pennisetum* species. *J Hered*. 88: 139-143.
- Naitani SP, Srivastava HC. 1965. Preliminary cytogenetical studies in *Psidium chinense*. *Current Science*. 18: 536-537.
- Ochatt SJ, Patat-Ochatt EM, Moessner A. 2011. Ploidy level determination within the context of *in vitro* breeding. *Plant Cell, Tissue and Organ Culture (PCTOC)*. 104:329–341.
- Ohri D, Kumar A. 1986. Nuclear DNA amounts in some tropical hardwoods. *Caryologia*. 39: 303-307.
- Otto FJ. 1990. DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. In: Darzynkiewicz Z, Crissman HA (eds) *Methods in cell biology*, vol 33. Academic Press, San Diego, pp 105-110.

- Pagliarini MS, Pereira MAS. 1992. Meiotic studies in *Pilocarpus pennatifolius* Lem. (Rutaceae). *Cytologia*. 57: 231-235
- Pagliarini MS. 1990. Meiotic behavior and pollen fertility in *Aptenia cordifolia* (Aizoaceae). *Caryologia*. 43: 157-162.
- Pires-Bione NC, Pagliarini MS, Ferraz de Toledo JF. 2000. Meiotic behavior of several Brazilian soybean varieties. *Genetics and Molecular Biology*. 23(3): 623-631.
- R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Raven PH. 1975. The bases of angiosperm phylogeny: cytology. *Annals of the Missouri Botanical Garden*. 62:724–764.
- Rayburn AL, Biradar DP, Nelson RL, McCloskey R, Yeater KM. 2004. Documenting Intraspecific Genome Size Variation in Soybean. *Crop Science*. 44: 261-264.
- Rojas-Rodríguez F, Torres-Córdoba G. 2013. Trees of the Central Valley of Costa Rica: reproduction Cas (*Psidium friedrichthalianum* (Berg) Ndz). *Kurú Mesoamerican Forest Magazine (Costa Rica)*. 120 (10): 1320-1323.
- Rojas-Garbanzo C, Wintera J, Montero ML, Zimmermann BF, Schiebert A. 2019. Characterization of phytochemicals in Costa Rican guava (*Psidium friedrichthalianum* Ndz) Fruit and stability of main compounds during juice processing - (U) HPLC-DAD- ESI-TQD-MS. *Journal of Food Composition and Analysis*. 75: 26-42.
- Soltis DE, Soltis PS, Bennett MD, Leitch IJ. 2003. Evolution of genome size in the Angiosperms. *Amer J Bot*. 90: 1596–1603.
- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. *Am J Bot*. 96 (1): 336-428.

- Soltis, P. S., Marchant, D. B., Van de Peer, Y., & Soltis, D. E. 2015. Polyploidy and genome evolution in plants. *Current opinion in genetics & development*. 35: 119-125.
- Souza AG, Resende LV, Lima IP, Martins LSS, Techio VH. 2015. Chromosome number and nuclear DNA amount in *Psidium* spp. resistant and susceptible to *Meloidogyne enterolobii* and its relation with compatibility between rootstocks and commercial varieties of guava tree. *Plant Systematics and Evolution*. 301: 231-237.
- Srivastava HC. 1977. Cytological studies in *Psidium friedrichsthalianum* Niedenzu. *Cytologia*. 42: 395-400.
- Torres C, Carvajal D, Rojas F, Arguedas M. 2011. Reproduction of tree and shrub species in the central region of Costa Rica. *Germinate*. (online). Cartago, CR, Institute of Technology of Costa Rica. Esc. Of Forest Engineer. Accessed: May 15, 2013.
- Vasconcelos LG, from Souza Santos E, from Oliveira Bernardes C, from Silva Ferreira MF, Ferreira A, Tuler CA, Macedo Carvalho JA, Fontes Pinheiro P, Praca-Fontes MM. 2019. Phytochemical analysis and effect of the essential oil of *Psidium* L. species on the initial development and mitotic activity of plants. *Environmental Science and Pollution Research*. 43: 1-13.
- Washburn JD, Birchler JA. 2014. Polyploids as a “model system” for the study of heterosis. *Plant Reprod*. 27:1–5.
- Wang K, Huang G, Zhu Y. 2016. Transposable elements play an important role during cotton genome evolution and fibre cell development. *Sci ChinaLifeSci*. 59: 112-121.
- Wendel JF, Jackson SA, Meyers BC, Wing RA. 2016. Evolution of plant genome architecture. *Genome Biol*. 17:37.
- Zhao D, Ferguson AA, Jiang N. 2016. What makes up plant genomes: the vanishing line between transposable elements and genes. *BBAGeneRegulMech*. 1859: 366-380.

## Tables

**Table 1.** Chromosomes number and genome size from different species of *Psidium* and *Eucalyptus* (Myrtaceae) determined in previous studies. The content of holoploid nuclear DNA (2C, pg DNA) and the content of monoploid DNA (1C, pg DNA) are also provided.

Species	2n	Ploidy level	Nuclear DNA content			
			2C (pg)	1C (pg)	1C (Mbp)*	Reference
<b>Genus <i>Eucalyptus</i></b>						
<i>Eucalyptus microcorys</i>	22	2x	1.040	0.520	508.56	Almeida-Carvalho <i>et al.</i> (2017)
<i>Eucalyptus botryoides</i>	22	2x	1.350	0.675	660.15	Almeida-Carvalho <i>et al.</i> (2017)
<b>Genus <i>Psidium</i></b>						
<i>Psidium guajava</i>	22	2x	0.950	0.475	464.55	Machado-Marques <i>et al.</i> (2016); Coser <i>et al.</i> (2012)
<i>Psidium guajava</i> (purple)	18	2x	0.990	0.495	484.11	Souza <i>et al.</i> (2015)
<i>Psidium guajava</i> ("Paluma")	22	2x	1.020	0.510	498.78	Souza <i>et al.</i> (2015)
<i>Psidium guajava</i> (white cultivar)	22	2x	0.507	0.253	247.43	Coser <i>et al.</i> (2012)
<i>Psidium guajava</i> (red cultivar)	22	2x	0.551	0.275	268.95	Coser <i>et al.</i> (2012)
<i>Psidium grandifolium</i> var. <i>cinereum</i>	44	4x	1.280	0.640	625.92	Costa and Forni-Martins (2009)
<i>Psidium grandifolium</i> var. <i>argenteum</i>	44	4x	0.820	0.410	400.98	Costa and Forni-Martins (2009)
<i>Psidium cattleianum</i>	44	4x	1.053	0.526	514.42	Costa and Forni-Martins (2006b)
<i>Psidium cattleianum</i>	44	4x	1.990	0.995	973.11	Souza <i>et al.</i> (2015)
<i>Psidium guineense</i>	44	4x	2.020	1.010	987.78	Souza <i>et al.</i> (2015)
<i>Psidium guineense</i>	44	4x	1.850	0.925	904.65	Machado-Marques <i>et al.</i> (2016)
<i>Psidium acutangulum</i>	44	4x	1.167	0.583	570.17	Costa and Forni-Martins (2009)

\*1pg DNA = 978 Mbp (Dolezel *et al.* 2003; Bennett *et al.* 2000).

**Table 2.** Collection sites of “Costa Rican Guava” used to determine genome size and chromosomal number. **PPT**: mean annual precipitation (mm); **Samples CMF**: number of seedlings used for flow cytometry; **Samples NM**: number of samples used to determine chromosomes count.

<b>Population name</b>	<b>Geographical coordinates</b>		<b>Altitude (m a.s.l.)</b>	<b>PPT (mm)</b>	<b>Temp (°C)</b>	<b>Samples CMF</b>	<b>Samples NM</b>
Cervantes	09°53'28.3"N	83°47'24.0"W	1465	2500	24	10	5
Guápiles	10°13'42.1"N	83°46'06.3"W	262	4535	27	10	5
Tacacorí	10°03'07.3"N	84°12'52.3"W	952	2100	23	10	5
Ciruelas	09°59'05.7"N	84°15'26.3"W	910	1900	23	10	5
Batán	10°04'34.4"N	83°22'37.2"W	114	3567	28	10	5
Escazú	09°55'08.3"N	84°07'42.6"W	2428	1929	24	10	5



**Table 3.** Parameters obtained by flow cytometry to determine the genome size of *Psidium friedrichsthalianum*. **NE**: Number of events obtained; **CV**: Coefficient of variation obtained; **2C (pg)**: holoploid nuclear DNA content obtained; 1pg DNA = 978 Mbp (Dolezel *et al.* 2003; Bennett *et al.* 2000).

Species	NE	CV	1C (pg)	2C (pg)	Mbp
<i>Psidium friedrichsthalianum</i>	2452 ± 0.001	2.95 ± 0.007	0.980 ± 0.005	1.960 ± 0.005	1916.88
<i>Glycine max (standard)</i>	2756 ± 0.005	3.01 ± 0.005			

## Figures

**Figure 1.** (a) Seedlings of *P. friedrichsthalianum* used to measure genome size by flow cytometry. (b and c) Flower buds of *P. friedrichsthalianum* used for cytogenetic observations.

**Figure 2.** Bivalent chromosomes of *P. friedrichsthalianum* in meiotic metaphase, stained with DAPI, scale bar 10um. (a, b and c) Samples from the populations of Cervantes, Tacacorí and Escazú respectively, showing 11 bivalent chromosomes. (d) Image showing chromosomes in trivalent, bivalent and univalent states.

**Figure 3.** Relative fluorescence intensity (propidium iodide (PI)) histogram obtained after a simultaneous cytometric analysis of nuclei of reference standard (*Glycine max*, 2C=2.50 pg of DNA) and *Psidium friedrichsthalianum* (2C=  $1.960 \pm 0.005$  pg).

**Figure 1.**



Figure 2.

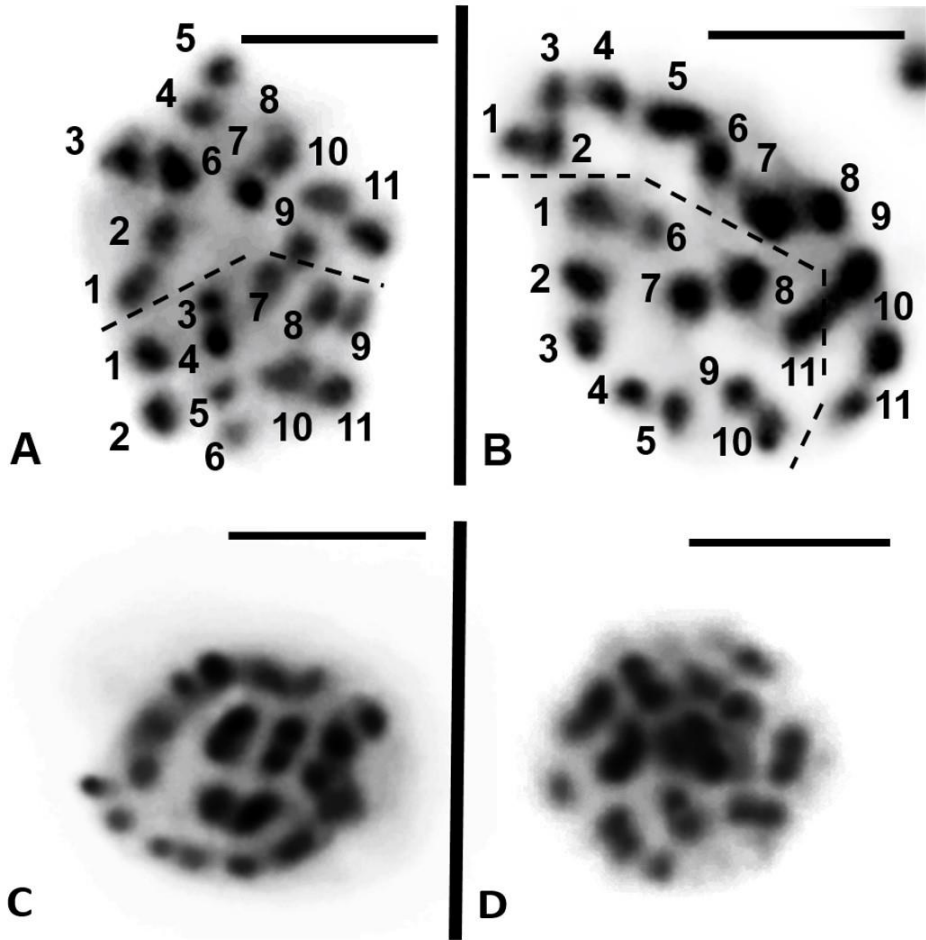
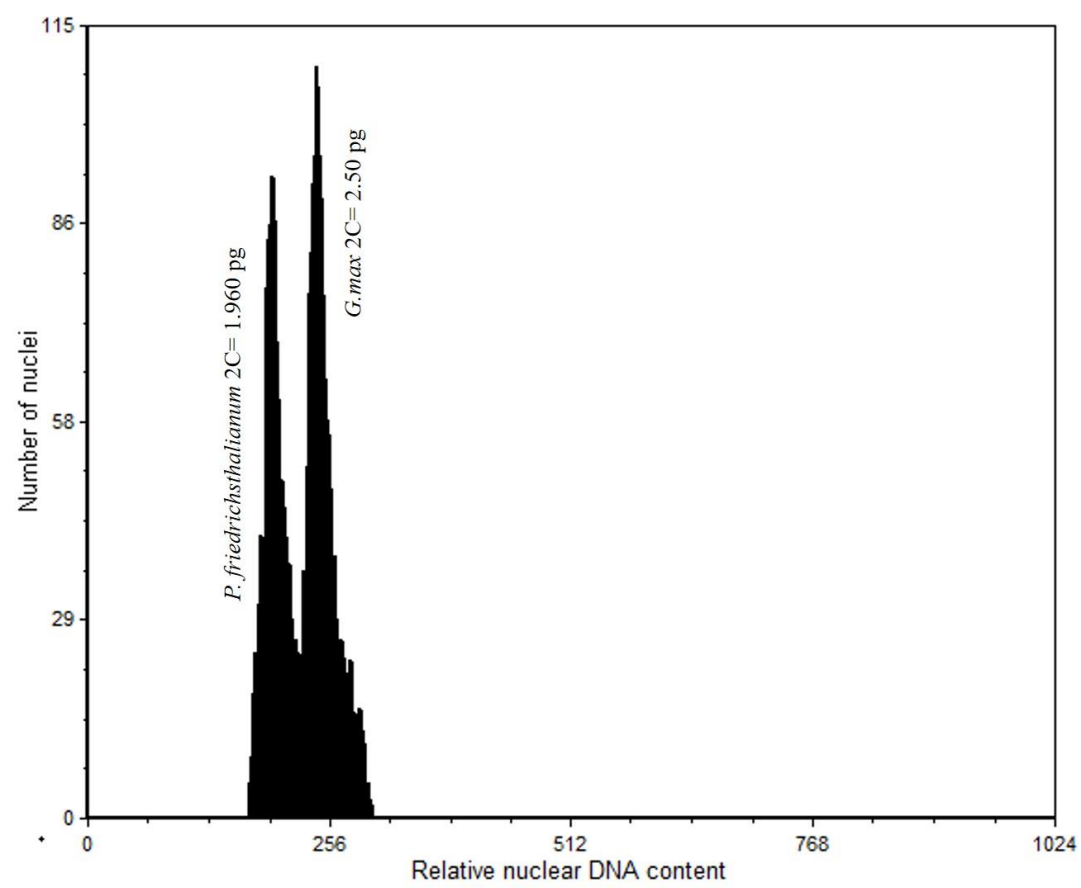


Figure 3.



**CAPÍTULO 2.** Caracterización de diversidad y estructura genética de poblaciones cultivadas de “Cas” (*Psidium friedrichsthalianum* Ndz) en Costa Rica utilizando ddRADseq<sup>1</sup>.

<sup>1</sup> Artículo con formato de revista PLOS ONE para ser sometido una vez sea traducido al inglés.

# Caracterización de diversidad y estructura genética de poblaciones cultivadas de “Cas” (*Psidium friedrichsthalianum* Ndz) en Costa Rica utilizando ddRADseq

## Resumen

El aumento de enfermedades crónicas y enfermedades no transmisibles (ENT) a nivel mundial ha impulsado el desarrollo de la "funcionalidad alimentaria". Un alimento funcional es aquel que además de sus componentes nutricionales, tiene metabolitos secundarios que afectan positivamente la salud más allá de los efectos nutricionales básicos, por ejemplo, reduciendo el riesgo de enfermedades. Diversas frutas nativas del continente americano han ganado atención como posibles alimentos funcionales, sin embargo, muchas de estas se encuentran aún en estado natural o “emergente” y no como cultivos comerciales; lo que provoca que las cosechas no tengan homogeneidad en contenido de nutrientes o concentración de metabolitos secundarios. Estas heterogeneidades reducen el potencial de comercializarlas a gran escala. En este contexto, *Psidium friedrichsthalianum* (O. Berg) Nied conocido popularmente como "Cas", es una especie cultivable de origen tropical cuyo centro de origen se presume es Costa Rica. Su fruto ha sido potencializado como “alimento funcional”, debido a su alto contenido polifenólico (proantocianidinas (PACs) tipo A y B) y a su gran capacidad antioxidante. Sin embargo, no se ha estudiado la diversidad y estructura genética de este cultivo y no existen bancos de germoplasma, lo cual limita el desarrollo comercial del cultivo e impide el desarrollo de programas de selección y mejoramiento genético. Por lo tanto, esta investigación tuvo como objetivo caracterizar diversidad y estructura genética en poblaciones cultivadas de Cas en Costa Rica utilizando ddRAD-seq, con el fin de identificar individuos o poblaciones con alta diversidad genética para su eventual conservación y/o selección en programas de mejoramiento genético. Como resultados obtuvimos 1 148 SNPs lo que sugiere que el método ddRAD-seq fue apto para genotipar las poblaciones cultivadas de Cas. Se obtuvo un valor promedio de heterocigosidad esperada ( $H_e$ ) igual a  $0.352 \pm 0.003$ , lo que sugiere alta diversidad genética en todas las poblaciones analizadas. Así mismo, hubo un exceso de heterocigotas en todas las poblaciones analizadas, posiblemente por causa de establecer estas plantaciones a partir de individuos o variedades altamente heterocigotas. Por otra parte, obtuvimos un  $F_{st}=0.220$  ( $p<0.05$ ), lo que evidencia estructura genética alta y significativa entre las poblaciones analizadas, además, esto sugiere que el flujo de genes en estas poblaciones es restringido y ocurre principalmente entre poblaciones cercanas o adyacentes a estas. Se realizó un dendrograma tipo Neighbor-Joining (NJ) y un análisis discriminante de componentes principales (DAPC), los cuales mostraron dos grupos de poblaciones, en congruencia con el análisis de STRUCTURE que propone  $K=2$ ; el primer grupo representado por la población de Escazú y el otro grupo distribuido casi homogéneamente entre las demás poblaciones. Nuestros resultados proponen que las poblaciones de Cas se estructuran principalmente por un factor de elevación y barreras naturales, pero no por distancia (IBD). Concluimos que el material genético altamente diverso observado en el presente estudio es promisorio para su resguardo o conservación; dicho material puede utilizarse eventualmente para el establecimiento de programas de selección y mejoramiento genético del cultivo de Cas.

**Palabras clave:** alimento funcional, Cas, diversidad genética, ddRAD-sequencing, secuenciación de próxima generación.

## Introducción

El aumento de enfermedades crónicas y enfermedades no transmisibles (ENT) a nivel mundial causadas por un estilo de vida sedentario y una dieta baja en nutrientes y antioxidantes, ha provocado que la ciencia y la industria agroalimentaria impulsen soluciones sostenibles como la “funcionalidad alimentaria” (Elmaliklis *et al.* 2019; Chang *et al.* 2018; McClements *et al.* 2009). Un alimento funcional es aquel que además de sus componentes nutricionales tiene metabolitos secundarios que afectan positivamente la salud, más allá de los efectos nutricionales básicos, reduciendo el riesgo de enfermedades (Farag *et al.* 2020; Bigliardi & Galati 2013; Binns & Howlett 2009). Diversas frutas originarias de América han ganado popularidad y atención como “alimentos funcionales”; algunos ejemplos son el Açaí (*Euterpe oleracea*), la Granadilla (*Passiflora ligularis*), la Uchuva (*Physalis peruviana*), y el Maqui (*Aristotelia chilensis*) (Romão *et al.* 2020; Etzbach *et al.* 2020; Fellah *et al.* 2019; Romero-González *et al.* 2019). Muchas de estas frutas se encuentran aún en estado natural o “emergente” y no como cultivos comerciales, provocando que las cosechas no tengan homogeneidad en contenido de nutrientes o concentración de metabolitos secundarios. Estas heterogeneidades reducen el potencial de comercializarlas a gran escala (Rufino *et al.* 2010; Gironés-Vilaplana *et al.* 2014).

*Psidium friedrichsthalianum* (O. Berg) Nied es una especie cultivable de origen tropical de la familia Myrtaceae, subfamilia Myrtoideae, tribu Myrteae (Lucas *et al.* 2019); cuyo origen se presume es en Costa Rica (Barahona and Rivera 1995). Produce un fruto conocido popularmente como “Cas”, “Guayaba ácida” o “Guayaba costarricense” (Barahona and Rivera 1995; Rojas-Rodríguez and Torres-Córdoba 2013) que es muy apetecido en la dieta de los costarricenses y centroamericanos (principalmente Nicaragua y El Salvador). En Costa Rica se vende en supermercados, “ferias del agricultor” y se



utiliza principalmente para hacer refrescos pasteurizados, postres, mermeladas y otros productos relacionados (León 2000). Por lo general, esta especie se cultiva en grupos pequeños, jardines caseros, potreros, y más recientemente, en plantaciones de mediano tamaño, pero, sin prácticas de manejo, apoyo técnico y promoción del cultivo (Rojas-Rodríguez and Torres-Córdoba 2013). Esta especie presenta individuos diploides ( $2n=2x=22$ ), con un tamaño de genoma de 1916 Mbp (capítulo 1), contiene flores hermafroditas, reproducción alógama y polinización principalmente realizada por insectos del orden Lepidóptera e Himenóptera (Barahona and Rivera 1995).

Algo destacable de esta especie es que su fruto ha sido potencializado como “alimento funcional”, por su alto contenido polifenólico (proantocianidinas (PACs) tipo A y B) y gran capacidad antioxidante (Rojas-Garbanzo *et al.*2019; Cuadrado-Silva *et al.*2017; Granados-Chinchilla *et al.*2016; Flores *et al.*2013; González *et al.*2011). Sin embargo, a pesar del potencial del Cas como “alimento funcional”, no hay información sobre la cantidad y distribución de su diversidad genética. Actualmente no existen bancos de germoplasma para esta especie, lo cual limita el desarrollo comercial del cultivo e impide el desarrollo de programas de selección y mejoramiento genético enfocados a cosechar frutos de Cas con un contenido alto y homogéneo de polifenoles.

Para el cultivo de Cas, la descripción de la diversidad y distribución genética es esencial para su conservación y reproducción, ya que permite identificar y resguardar material genético altamente diverso que puede ser utilizado para reproducir nuevos cultivares, favorecer la domesticación del cultivo o bien desarrollar nuevas variedades según las necesidades productivas (Olsen *et al.* 2013). Aunado a esto, los marcadores moleculares han sido utilizados ampliamente en estudios de diversidad genética en plantas (Kumar and Agrawal 2017; Treuren and Hintum 2014; Bordoni and Hodgkin 2012); un ejemplo de marcadores moleculares son los polimorfismos de un solo

nucleótido (SNPs) (Thavamanikumar *et al.* 2011). Los SNPs se caracterizan por ser abundantes en el genoma en comparación a otros marcadores, además una gran mayoría de SNPs se encuentran en regiones genómicas codificantes, tienen menor tasa de error y mayor rentabilidad en comparación con los microsatélites (SSR) y AFLPs (Mammadov *et al.* 2012; Kumar *et al.* 2017). Así mismo, los SNPs poseen una baja tasa de mutación ( $\sim 10^{-8}$ ), siendo más adecuados para marcadores de selección a largo plazo (McKay *et al.* 2008; Lin *et al.* 2010).

Con la aparición de la secuenciación de próxima generación (NGS, con sus siglas en inglés) se han implementado métodos para el desarrollo y genotipado de miles de marcadores en todo el genoma de las plantas (Bao *et al.* 2020; Basak *et al.* 2019; Davey *et al.* 2011). La secuenciación de ADN asociada a sitios de restricción (ddRADseq) (Li *et al.* 2019; Sudan *et al.* 2019; Bai *et al.* 2018; Davey *et al.* 2011), es una técnica que genera una gran cantidad de marcadores moleculares con polimorfismos de un solo nucleótido (SNPs) a partir de una sub-representación del genoma (Peterson *et al.* 2012). A diferencia de la secuenciación del genoma completo, en ddRADseq solo se dirige a un subconjunto del genoma, pero con la ventaja de que no requiere ningún conocimiento previo del genoma de la especie de interés, proporcionando buen rendimiento y rentabilidad para estudios de diversidad genética (Guillardín-Calvo *et al.* 2019; Pyne *et al.* 2017).

Esta investigación tuvo como objetivo caracterizar la diversidad y estructura genética en poblaciones cultivadas de Cas (*Psidium friedrichsthalianum* Ndz) en Costa Rica utilizando el método ddRAD-seq, con el fin de identificar individuos o poblaciones con alta diversidad para su eventual conservación y/o selección en programas de mejoramiento genético.

## **Materiales y métodos**

### ***Colecta de material vegetal***

Se muestrearon 75 individuos de cinco poblaciones cultivadas de Cas de Costa Rica (Fig 1). El cuadro 1 muestra las coordenadas geográficas, elevación (msnm), precipitación promedio anual (mm), temperatura promedio (°C) y número de individuos analizados por población. Las poblaciones eran plantaciones de distintas regiones del país que se encontraban distribuidas en diferentes elevaciones desde los 114 msnm hasta los 1865 msnm. Siempre se tomaron muestras foliares de árboles reproductivos y se tuvo cuidado de recolectar muestras de individuos que estaban separados por al menos 10 metros entre sí para evitar la colecta de genets. De cada individuo se recolectaron hojas jóvenes y se colocaron en tubos con sílica gel y posteriormente se transportaron dentro de una hielera hasta el laboratorio para extracción de ADN.

### ***Secuenciación con el método ddRAD-seq***

Se extrajo ADN genómico de todos los individuos con el protocolo de extracción basado en CTAB como describe Doyle & Doyle (1987). Se cuantificó el ADN mediante NanoDrop<sup>TM</sup>2000 (Thermo Scientific<sup>TM</sup>) y se verificó que las muestras tuvieran una relación de absorbancia 260/280 entre 1.8 a 2.0, además, se observó la integridad del ADN en un gel de agarosa al 1%. Posteriormente, las muestras de ADN se enviaron a Texas A&M University Corpus Christi para realizar la secuenciación con el método ddRAD-seq, siguiendo el protocolo que describe Peterson *et al.* (2012).

### ***Análisis de secuencias (lecturas) e identificación de SNPs***

Primeramente, se realizó el de-multiplexado de las lecturas, y se analizó la calidad de las mismas mediante Fastqc v0.11.5 (Andrews 2010). Se realizó un filtrado de las lecturas que presentaron baja calidad de bases (filtering low-quality bases), utilizando Trimmomatic-0.36 (Bolger 2014) con los siguientes parámetros: SLIDINGWINDOW:4:30 y MINLEN:100. Las lecturas remanentes se alinearon contra el genoma de *Psidium guajava* (4728 contigs, con un total de 0.38 GB, versión 1.0), utilizando el alineador BOWTIE2 v2.2.9 con los parámetros estándar del programa (Langmead *et al.* 2012). Luego, se utilizó SAMtools v1.9 (Li *et al.* 2009) para convertir los archivos del alineamiento de formato SAM a formato BAM con el fin de visualizarlos en el programa Integrate Genomics Viewer (IGV) (© 2013-2018 Broad Institute and the Regents of the University of California). Enseguida, a cada archivo BAM se le añadió una etiqueta RG, utilizando PICARD v2.18.9 y se realizó un ordenamiento con SAMtools v1.9. Posteriormente, se realizó el SNPs variant calling con Freebayes (Garrison and Marth 2013), utilizando dos opciones: --min-coverage 15 (cobertura mínima de 15X) y --i (excluyendo indels). Enseguida, los SNPs llamados se filtraron utilizando VCFtools v0.1.15 (Danecek *et al.* 2011), contra los siguientes umbrales: minor allele frequency (MAF)  $\geq 0.05$  (se incluyeron sitios con una frecuencia de alelo mayor o igual que 0.05), --max-alleles 2 (se incluyeron solamente sitios bialélicos), --max-missing 0.8 (se excluyeron sitios en función de la proporción de datos faltantes (definido entre 0 y 1, donde 0 permite sitios que faltan por completo y 1 indica que no se permiten datos faltantes), --minQ 30 (se incluyeron sitios con un valor de calidad por encima de este umbral), --minDP 5 (se incluyeron sitios con coberturas mayores o iguales a 5X), --hwe 0.05 (se excluyeron sitios que fallan prueba de equilibrio de Hardy-Weinberg (HWE) a

$p < 0.05$ ). En la Figura 2 se observa en resumen el *pipeline* bioinformático utilizado en el presente estudio para el análisis de las lecturas ddRAD-seq.

### ***Estimativas básicas de diversidad y estructura genética***

Se calcularon estadísticas básicas de diversidad genética; heterocigosidad observada ( $H_o$ ), heterocigosidad esperada ( $H_E$ ), número de alelos privados y estadísticas F de Wright: coeficiente de endogamia (FIS) y el estadístico FST (Weir and Cockerham 1984). Además, se realizó un dendrograma tipo neighbor-joining (NJ) con las distancias genéticas de Nei (Saitou and Nei 1987) y un análisis discriminante de componentes principales (DAPC) (Jombart *et al.* 2010) para encontrar alguna estructura subyacente en las poblaciones. Lo anterior se calculó utilizando los paquetes “poppr” (Kamvar *et al.* 2014), “adegenet” (Jombart 2008), “hierfstat” (Goudet 2005) y “pegas” (Paradis 2010) en el lenguaje estadístico R (R Core Team 2018). Así mismo, se realizó un análisis de agrupamiento bayesiano implementado por STRUCTURE ver. 2.3.4 (Pritchard *et al.* 2000), para estimar el número K de grupos representados por el conjunto de datos. Se realizaron veinte ejecuciones estableciendo el número de agrupaciones (K) de 2 a 6. Cada ejecución consistió en un quemado de 25 000 pasos de la cadena de Markov (MCMC), seguidos de 50 mil réplicas. La elección del número más probable de grupos (K) se llevó a cabo mediante el cálculo de las estadísticas  $\Delta K$ , que se basan en la tasa de cambio en la probabilidad de registro de los datos entre los valores de K sucesivos, según lo descrito por Evanno *et al.* (2005) utilizando Structure Harvester (Earl & vonHoldt 2012). Además, se realizó una prueba de Mantel de distancias genéticas y geográficas utilizando la función “mantel.randtest” en el programa R (Mantel 1967), con el fin de comprobar aislamiento por distancias entre las poblaciones.

## Resultados

### *Análisis de secuencias (lecturas) e identificación de SNPs*

La secuenciación ddRAD generó un promedio de 5 598 710 (SD 402.675) de lecturas por individuo (forward y reverse). Después del procesamiento con Trimmomatic-0.36, en promedio quedaron 1 925 703 (SD 92.277) de lecturas forward y 685 205 (SD 34.789) lecturas reverse por individuo (Fig S3); las lecturas remanentes presentaron una calidad promedio de Phred score de 30 (Q score) y un contenido promedio de 0.444 GC.

El 54% (102 921 993 lecturas) de las lecturas se alinearon contra el genoma de *Psidium guajava* (4 728 contigs, con un total de 0.38 GB). Además, el buscador de variantes (“variant calling”) identificó un total de 1 168 570 de SNPs. Con el primer filtro del programa *VCFtools* se redujo a 393 329 SNPs; el segundo filtro a 161 468 SNPs y el tercer filtro concluyó con 1 148 SNPs en total. Los SNPs se clasificaron según la sustitución de nucleótidos como transiciones (C↔T o A ↔G) o transversiones (A↔C, C↔G, A↔T, G↔T) y se observó una mayor frecuencia de transiciones (57.46%) que de transversiones (42.54%) (Fig S4).

### *Estadísticas de diversidad genética*

Los valores de heterocigosidad observada ( $H_o$ ) y heterocigosidad esperada ( $H_E$ ) se mostraron similares en todas las poblaciones, ambas con un promedio de  $H_o = 0.475 \pm 0.003$  y  $H_E = 0.352 \pm 0.003$ ; así mismo los valores de heterocigosidad observada ( $H_o$ ) fueron mayores a los valores de heterocigosidad esperada ( $H_E$ ) (Cuadro 2), evidenciando un exceso de heterocigotas en todas las poblaciones.

### ***Estructura genética***

Se encontró diferencias significativas en las frecuencias alélicas de las poblaciones ( $F_{ST} = 0.220$ ;  $p < 0.05$ ) (Cuadro 2). No existe una correlación significativa entre la distancia geográfica y genética entre poblaciones ( $r = 0.453$ ;  $p\text{-value} = 0.097$ ) (Fig S5). Las distancias genéticas de Nei de todas las poblaciones se observan en el Cuadro 3, las poblaciones Ciruelas y Tacacorí mostraron la menor distancia genética ( $D_{Nei} = 0.00013$ ); mientras que las poblaciones Ciruelas y Escazú presentaron la mayor distancia genética ( $D_{Nei} = 0.00837$ ).

El dendrograma tipo Neighbor-Joining (NJ) mostró dos agrupamientos claros, uno conformado por las poblaciones Cervantes, Tacacorí y Ciruelas, y un segundo agrupamiento conformado por las poblaciones Batán y Escazú (Fig 3). El análisis discriminante de componentes principales (DAPC), también agrupa las poblaciones Cervantes, Tacacorí y Ciruelas; pero separa las poblaciones de Batán y Escazú (Fig 4). El análisis de STRUCTURE indicó que el número K de conglomerados óptimo es  $K=2$ . El primer grupo representado principalmente por la población de Escazú (Fig 5) y el otro grupo distribuido casi homogéneamente entre las demás poblaciones.

### **Discusión**

El alto porcentaje de alineamiento de las secuencias de Cas contra el genoma del congénere *Psidium guajava* (Guayaba) sugiere que el uso de este genoma fue adecuado como genoma de referencia para realizar los análisis del presente estudio. De la misma manera, sugiere que ambas especies (Cas y Guayaba) no presentan aún mucha divergencia (Peterson *et al.* 2012).

Nuestros resultados sugieren que las poblaciones de Cas presentan alta diversidad genética ( $H_E=0.352$ ). Una alta diversidad genética en poblaciones de Cas puede deberse

en parte a su sistema de reproducción. Se ha demostrado que las especies de plantas con polinización cruzada como el Cas, presentan mayor variación genética dentro de sus poblaciones, en comparación con especies autógamas o especies con sistemas de apareamiento mixto (Loveless & Hamrick 1984; Hamrick & Godt 1996). Por lo general, la polinización cruzada promueve el flujo génico y el movimiento de genes entre poblaciones, aumentando la diversidad genética (Booy *et al.* 2000; Loveless & Hamrick 1984; Hamrick *et al.* 1992; El-Bakry *et al.* 2014). Así mismo, una alta diversidad genética en Cas puede deberse a su amplia distribución en el país y la existencia de poblaciones naturales que mantienen el acervo genético original de la especie (Loveless & Hamrick 1984; Hamrick & Godt 1996). De la misma manera, es esperable observar alta diversidad genética en Cas debido a que este cultivo no ha pasado por procesos de domesticación extensos, en comparación con otros cultivos como cacao (Ji *et al.* 2013), arroz (Li 2015) y soya (Li *et al.* 2010).

Una estructura genética significativa en el presente estudio indica que el flujo de genes entre las poblaciones de Cas se encuentra restringido; y ocurre principalmente entre poblaciones cercanas o adyacentes a estas. Nuestros resultados proponen que las poblaciones de Cas se estructuran principalmente por un factor de elevación y barreras naturales, pero no por distancia (IBD) (Fig S5). Los paisajes montañosos con diferentes elevaciones limitan el flujo de genes entre poblaciones de plantas (Cain *et al.* 2000; Theurillat and Guisan 2001), causando diferenciación genética entre ellas, en comparación a especies que se encuentran limitadas a un único plano elevacional (Till-Bottraud and Gaudeul 2002). Nuestros resultados consistentemente separan la población Escazú de las otras; probablemente por su elevación (1865 msnm) que difiere marcadamente de las demás poblaciones (Cuadro 1). Es posible que estas diferencias en elevación limiten el flujo de genes en *P. friedrichsthalianum* (Cas), causando diferencias



en frecuencias alélicas entre poblaciones por deriva. Así mismo, el agrupamiento entre las poblaciones Tacacorí, Ciruelas y Cervantes probablemente se deba a que tienen un mismo origen genético y provienen de un mismo cultivar, aunado a que presentan elevaciones similares que pueden mantener el flujo génico entre ellas (Fig. 4).

Así mismo, el patrón de estructura génica observado puede deberse también al trasiego de material (semillas o plántulas) entre fincas aledañas, lo cual es una práctica muy común en sistemas de cultivo para aumentar la productividad (Sousa *et al.* 2017; Kawa *et al.* 2013). La forma en que los agricultores cultivan el Cas puede ser otra causa de estructuración genética en estas poblaciones. Usualmente los agricultores de Cas cosechan sus frutos y obtienen de ellos las semillas que utilizan para producir más individuos en sus fincas, esto por lo general provoca endogamia y aumenta la diferenciación genética entre las poblaciones (Ellegren and Galtier 2016).

El exceso de heterocigotas observado en las poblaciones de Cas puede deberse a la selección de individuos más heterocigotos cuando se establecieron las plantaciones inicialmente (selección sobredominante) (Nei 1987; Ohta and Kimura 1970). Por lo general, el establecimiento de cultivos conlleva la selección de individuos con alta heterosis que reflejan mejor adaptación, ventajas de rendimiento y producción (mejor calidad de fruto, mayor crecimiento, mayor resistencia a enfermedades, etc.) (Coulso *et al.* 1998). Por lo tanto, sugerimos que un posible establecimiento de plantaciones de Cas a partir de pocos individuos altamente heterocigotas, aunado con un trasiego de plantas entre productores, explica los altos niveles de heterocigosidad observada en estos cultivares. Por otra parte, Hartl & Clark (1997) y Glémin *et al.* (2001) citan los sistemas de auto-incompatibilidad como posible causa de exceso de heterocigotas en plantas, sin embargo, la auto-incompatibilidad en la especie de Cas no es clara aún, por lo tanto, para

comprobar dicha hipótesis recomendamos hacer más estudios relacionados. Otra posible explicación del exceso de heterocigotas en el presente estudio podría deberse a la presencia de secuencias parálogas en el genoma de Cas (Verdu *et al.* 2016). Si bien existen muchas ventajas al usar el método ddRAD-seq para genotipado, una desventaja es que las secuencias parálogas son difíciles de separar de sus loci, debido a que se obtienen secuencias muy cortas y con gran similitud entre ellas (Davey *et al.* 2013). Diferentes autores recomiendan filtrar estas secuencias, sin embargo, esto se puede dificultar si no existe un genoma de referencia (Baird *et al.* 2008; Miller *et al.* 2007).

## **Conclusiones**

El presente estudio evidencia alta diversidad genética en las poblaciones cultivadas de Cas en Costa Rica, sugerimos que esto puede deberse a la amplia distribución del Cas en el país, al sistema de reproducción cruzada que presenta el Cas, y la existencia de poblaciones naturales que mantienen el acervo genético original de la especie (Loveless & Hamrick 1984; Hamrick & Godt 1996). De la misma manera, es esperable observar alta diversidad genética en Cas debido a que este cultivo no ha pasado por procesos de domesticación más extensos, en comparación con otros cultivos (Ji *et al.* 2013). Observamos que las diferencias de elevación entre las poblaciones analizadas parece ser un factor importante de estructuración genética en Cas, también, el trasiego de material (semillas o plántulas) entre fincas aledañas y el método de cultivo del Cas son posibles causantes de la estructuración genética observada. Proponemos que el exceso de heterocigotas observado en el presente estudio, puede deberse principalmente al establecimiento de las plantaciones de Cas a partir de grupos de individuos o variedades altamente heterocigotas.

Finalmente, el material genético altamente diverso observado en el presente estudio es promisorio para su resguardo o conservación; dicho material puede utilizarse eventualmente para el establecimiento de programas de selección y mejoramiento genético del cultivo de Cas. Así mismo, es de importancia complementar dicha información con datos de perfiles de contenido de antioxidantes o con descriptores morfológicos (tamaño del fruto) de cada población analizada, para establecer la verdadera relación entre variantes genotípicas y fenotípicas, y seleccionar material genético altamente productivo con el cual se pueda establecer cultivos comerciales con un valor agregado de alimento funcional.

## Referencias

- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Bai B, Wang L, Zhang YJ. 2018. Developing genome-wide SNPs and constructing an ultrahigh-density linkage map in oil palm. *Sci Rep* 8, 691.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*. 3: 1–14.
- Bao W, Ao D, Wuyun, T. 2020. Development of 85 SNP markers for the endangered plant species *Prunus mira* (Rosaceae) based on restriction site-associated DNA sequencing (RAD-seq). *Conservation Genet Resour*.
- Baraona M, Rivera G. 1995. Development of jocote (*Spondias purpurea* L.) and cas (*Psidium friedrichsthalianum* Ndz) in the premontano humid forest of Costa Rica. *Mesoamerican Agronomy*. 6: 23-31.
- Basak M, Uzun B, Yol E. 2019. Genetic diversity and population structure of the Mediterranean sesame core collection with use of genome-wide SNPs developed by double digest RAD-Seq. *PLoS ONE* 14(10): e0223757.
- Bigliardi B, Galati F. 2013. Innovation trends in the food industry: the case of functional foods. *Trends in*

- Food Science & Technology. 31:118–129.
- Binns N, Howlett J. 2009. Functional foods in Europe: international developments in science and health claims. *European Journal of Nutrition*. 48: S3–S13.
- Bolger AM. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114–2120.
- Booy G, Hendriks RJJ, Smulders MJM, Van Groenendael JM, Vosman B. 2000. Genetic diversity and the survival of populations. *Plant Biology*. 2:79–395.
- Bordoni P, Hodgkin T. 2012. Climate change and the conservation of plant genetic resources. *J Crop Improv*. 26(3):329–45.
- Brunet J, Holmquist KGA. 2009. The influence of distinct pollinators on female and male reproductive success in the rocky mountain columbine. *Molecular Ecology*. 18: 3745 – 3758.
- Cain ML, Milligan BG, Strand AE. 2000. Long-distance dispersal in plant populations. *American Journal of Botany*. 87:1217–1227.
- Chang SK, Alasalvar C, Shahidi F. 2018. Superfruits: Phytochemicals, antioxidant efficacies, and health effects – A comprehensive review, *Critical Reviews in Food Science and Nutrition*.
- Coulson TN, Pemberton JM, Albon SD. 1998. Microsatellites reveal heterosis in red deer. *Proceedings of the Royal Society of London. Series B, Biological Sciences*. 265: 489–495.
- Cuadrado-Silva CT, Pozo-Bayón MÁ, Osorio C. 2017. Targeted Metabolomic Analysis of Polyphenols with Antioxidant Activity in Sour Guava (*Psidium friedrichsthalianum* Ndz.). *Fruit Molecules*. 22 (1): 11.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST. 2011. The variant call format and VCFtools. *Bioinformatics*. 27(15):2156–8.
- Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. 2011. Special features of RAD sequencing data: Implications for genotyping. *Molecular Ecology*. 22: 3151–3164.
- Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. 2013. Special features of RAD sequencing data: implications for genotyping. *Molecular Ecology*. 22: 3151–3164.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochemical Bulletin*. 19: 11–15.

- Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserve Genet Resour* 4:359-361.
- El-Bakry A, Hammad IA, Rafat FA. 2014. Polymorphism in *Calotropis procera*: preliminary genetic variation in plants from different phytogeographical regions of Egypt. *Rendiconti Lincei*. 25: 471–477.
- Ellegren H, Galtier N. 2016. Determinants of genetic diversity. *Nat Rev Genet*.17:422–433.
- Elmaliklis IN, Liveri A, Ntelis B, Paraskeva K, Goulis I, Koutelidakis AE. 2019. Increased Functional Foods' Consumption and Mediterranean Diet Adherence May Have a Protective Effect in the Appearance of Gastrointestinal Diseases: A Case–Control Study. *Medicines*. 6: 50.
- Etzbach L, Meinert M, Faber T, Klein C, Schieber A, Weber F. 2020. Effects of carrier agents on powder properties, stability of carotenoids, and encapsulation efficiency of goldenberry (*Physalis peruviana* L.) powder produced by co-current spray drying. *Current Research in Food Science*. 24:3078.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14: 2611–2620.
- Farag MA, Abdelwareth A, Sallam IE, el Shorbagi M, Jehmlich N, Fritz-Wallace K, Schäpe S, Rolle-Kampczyk U, Ehrlich A, Wess LA, von Bergen M. 2020. Metabolomics reveals impact of seven functional foods on metabolic pathways in a gut microbiota model. *Journal of Advanced Research*. 23: 47–59.
- Fellah B, Rocchetti G, Senizza B, Giuberti G, Bannour M, Ferchichi A, Lucini L. 2019. Untargeted metabolomics reveals changes in phenolic profile following in vitro large intestine fermentation of non-edible parts of *Punica granatum* L. *Food Research International*. 39: 45-65.
- Flores G, Dastmalchi K, Wu SB, Whalen K, Dabo AJ, Reynertson K. 2013. Phenol-rich extract of the Costa Rican guava pulp (*Psidium friedrichsthalianum* Ndz) with antioxidant and anti-inflammatory activity. Potential for COPD therapy. *Food Chemistry*. 14: 889-895.
- Garrison E, Marth G. 2013. Haplotype-based variant detection from short-read sequencing arXiv: 1207.3907.
- Gironés-Vilaplana A, Baenas N, Villano D, Speisky H, Garcia-Viguera D, Moreno DA. 2014. Evaluation

- of Latin-American fruits rich in phytochemicals with biological effects. *Journal of Functional Foods*. 7:599–608.
- Glémin S, Bataillon T, Ronfort J, Mignot A, Olivieri I. 2001. Inbreeding depression in small populations of self-incompatible plants. *Genetics*. 159: 1217–1229.
- González E, Vaillant F, Pérez AG, Rojas T. 2011. Antioxidant protection mediated by in vitro cells of human erythrocytes by some common tropical fruits. *J Nutr Food Sci*. 2: 139.
- Goudet J. 2005. Hierfstat, a package for R to compute and test variance components and F statistics. *Molecular Ecology Notes*. 5:184-186.
- Granados-Chinchilla F, Villegas E, Molina A, Arias C. 2016. Composition, Chemical Fingerprinting and Antimicrobial Assessment of Costa Rican Cultivated Guavas (*Psidium friedrichsthalianum* (O. Berg) Ndz and *Psidium guajava* L.) Essential Oils from Leaves and Fruits. *Nat Prod Chem Res*. 4:5.
- Guillardín-Calvo L, Mora-Márquez F, Soto Á. 2019. RADDESIGNER: a workflow to select the optimal sequencing methodology in genotyping experiments on woody plant species. *Tree Genetics & Genomes*. 15: 64.
- Hamrick JL, Godt MJ. 1996. Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences* 351:1291–1298.
- Hamrick JL, Godt MJW, Sherman-Broyles SL. 1992. Factors influencing levels of genetic diversity in woody plant species. *New Forest*. 6: 95-124.
- Hartl D, Clark A. 1997. *Principles of Population Genetics*, 3<sup>rd</sup> ed. Sinauer Associates, Sunderland, Massachusetts.
- Ji K, Zhang D, Motilal, LA. 2013. Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. *Genet Resour Crop Evol*. 60: 441–453.
- Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11:94.
- Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*.

24:1403–1405.

- Kamvar ZN, Tabima JF, Grünwald NJ. 2014. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*. 2:28-45.
- Kawa NC, McCarty C, Clement CR. 2013. Manioc varietal diversity, social networks, and distribution constraints in rural Amazonia. *Curr. Anthropol*. 54: 764-770.
- Kumar J, Agrawal V. 2017. Analysis of genetic diversity and population genetic structure in *Simarouba glauca* DC. (an important bio-energy crop) employing ISSR and SRAP markers Jatin Kumar, Veena Agrawal. *Industrial Crops and Products* 100: 198–207.
- Kumar S, Banks TW, Cloutier S. 2017. SNP discovery through next-generation sequencing and its applications. *Int J Plant Genomics*. 4:37-48.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 9(4):357-9.
- Lee JH, Natarajan S, Biswas MK, Shirasawa K, Isobe S, Kim HT. 2018. SNP discovery of Korean short day onion inbred lines using double digest restriction site-associated DNA sequencing. *PLoS ONE* 13(8): e0201229
- León J. 2000. Botany of tropical crops. San José, Costa Rica: Ed. Ilca. pp. 445.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25:2078–2079.
- Li H. 2015. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 27: 2987–2993.
- Li S, Lv S, Yu K. 2019. Construction of a high-density genetic map of tree peony (*Paeonia suffruticosa* Andr. Moutan) using restriction site associated DNA sequencing (RADseq) approach. *Tree Genetics & Genomes* 15, 63.
- Li YH, Zhang WLC, Li-Juan Qiu. 2010. Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. *New Phytologist Trust*. 20:56-89.
- Lin BZ, Sasazaki S, Mannen H. 2010. Genetic diversity and structure in *Bos taurus* and *Bos indicus* populations analyzed by SNP markers. *Anim. Sci. J.* 81 281–289.

- Loveless MD, Hamrick JL. 1984. Ecological determinants of genetic structure in plant populations. *Ann Rev Em Syst.* 15:65-95.
- Lucas EJ, Holst B, Sobral M, Mazine FF, Nic Lughadha EM, Barnes Proença CE, Ribeiro da Costa I, Vasconcelos TNC. 2019. A New Subtribal Classification of Tribe Myrteae (Myrtaceae). *Systematic Botany.* 44(3): pp. 560–569.
- Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S. 2012. SNP markers and their impact on plant breeding. *Int J Plant Genomics.* (30): 23-36.
- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research.* 27:209-220.
- McClements DJ, Decker EA, Park Y, Weiss J. 2009. Structural Design Principles for Delivery of Bioactive Components in Nutraceuticals and Functional Foods, *Critical Reviews in Food Science and Nutrition.* 49:6, 577-606
- McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W. 2008. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. *BMC Genetics* 9:37.
- Miller M, Dunham J, Amores A, Cresko W, Johnson E. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17: 240–248.
- Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583-590.
- Ohta T, Kimura M. 1970. Development of associative overdominance through linkage disequilibrium in finite populations. *Genetical Research.* 16. 165–177.
- Olsen KM, Wendel JF. 2013. A Bountiful Harvest: Genomic Insights into Crop Domestication Phenotypes. *Annu Rev Plant. Biol.* 64: 47–70.
- Paradis E. 2010. pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics.* 26: 419–420.
- Penjor T, Mimura T, Kotoda N, Matsumoto R, Nagano JA, Honjo NM. 2016. RAD-Seq analysis of typical and minor Citrus accessions, including Bhutanese varieties. *Breed Sci.* 66(5):797–807.



- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*.7: 37-56.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155: 945–959.
- Pyne R, Honig J, Vaiciunas J, Koroch A, Wyenandt C, Bonos S, Simon J. 2017. A first linkage map and downy mildew resistance QTL discovery for sweet basil (*Ocimum basilicum*) facilitated by double digestion restriction site associated DNA sequencing (ddRADseq). *PloS one*. 12(9).
- R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Rojas-Garbanzo C, Wintera J, Montero ML, Zimmermann BF, Schieber A. 2019. Characterization of phytochemicals in Costa Rican guava (*Psidium friedrichsthalianum* Ndz) Fruit and stability of main compounds during juice processing - (U) HPLC-DAD- ESI-TQD-MS. *Journal of Food Composition and Analysis*. 75: 26-42.
- Rojas-Rodríguez F, Torres-Córdoba G. 2013. Trees of the Central Valley of Costa Rica: reproduction Cas (*Psidium friedrichsthalianum* (Berg) Ndz). *Kurú Mesoamerican Forest Magazine (Costa Rica)*. 120 (10): 1320-1323.
- Romão M, de Bem GF, Santos IB, Soares R, Ognibene D, de Moura RS, da Costa CA, Resende A. 2020. Açaí (*Euterpe oleracea* Mart.) seed extract protects against hepatic steatosis and fibrosis in high-fat diet-fed mice: Role of local renin-angiotensin system, oxidative stress and inflammation. *Journal of Functional Foods* 65: 103-726.
- Romero-González J, Shun Ah-Hen K, Lemus-Mondaca R, Muñoz-Fariña O. 2019. Total phenolics, anthocyanin profile and antioxidant activity of maqui, *Aristotelia chilensis* (Mol.) Stuntz, berries extract in freeze-dried polysaccharides microcapsules. *Food Chemistry*. (19):322-345.
- Rufino MDS, Alves MRE, de Brito ES, Perez-Jimenez P, Calixto S, Mancini-Filho J. 2010. Bioactive compounds and antioxidant capacities of 18 non-traditional tropical fruits from Brazil. *Food Chemistry*. 121:996–1002.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees.

- Satya P, Banerjee R, Karan M, Mukhopadhyay E, Chaudhary B, Bera A, Maruthi RT, Sarkar SK. 2016. Insight into genetic relation and diversity of cultivated and semi-domesticated under-utilized *Crotalaria* species gained using start codon targeted (SCoT) markers. *Biochemical Systematics and Ecology*. 66: 24-32.
- Shirasawa K, Hirakawa H, Isobe S. 2016. Analytical work flow of double-digest restriction site-associated DNA sequencing based on empirical and in silico optimization in tomato. 23(2):145–153.
- Shirasawa K, Tanaka M, Takahata Y, Ma D, Cao Q, Liu Q. 2017. A high-density SNP genetic map consisting of a complete set of homologous groups in autohexaploid sweetpotato (*Ipomoea batatas*). *SciRep*. 7:44207.
- Singh N, Choudhury DR, Singh AK, Kumar S, Srinivasan K, Tyagi RK, Singh NK, Singh R. 2013. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PloS One*.8:12.
- Sousa S.B, Silva G.F, Dias M.C, Clement C.R, Sousa N.R. 2017. Farmer variety exchange along Amazonian rivers influences the genetic structure of manioc maintained in a regional Brazilian Gene Bank. *Genet Mol Res*. 16(3):45-78.
- Sudan J, Singh R, Sharma S. 2019. ddRAD sequencing-based identification of inter-genepool SNPs and association analysis in *Brassica juncea*. *BMC Plant Biol* 19, 594.
- Thavamanikumar S, McManus LJ, Tibbits JFG, Bossinger G. 2011. The significance of single nucleotide polymorphisms (SNPs) in *Eucalyptus globulus* breeding programs. *Australian Forestry*. 74:23-29.
- Theurillat JP, Guisan A. 2001. Potential impact of climate change on vegetation in the European Alps: a review. *Climate Change*. 50: 77-109.
- Till-Bottraud I, Gaudeul M, Körner C, Spehn EM. 2002. Intraspecific genetic diversity in alpine plants. *Mountain biodiversity: a global assessment*. New York, NY Parthenon Publishing. pg. 23-34.
- Treuren R, Hintum TJL. 2014. Next-generation genebanking: plant genetic resources management and utilization in the sequencing era. *Plant Genet Resour*.12:298–307.
- Verdu C, Guichoux E, Quevauvillers S, De Thier O, Laizet Y. 2016. Dealing with paralogy in RADseq

data: in silico detection and single nucleotide polymorphism validation in *Robinia pseudoacacia* L. *Ecol Evol.* 6(20): 7323–7333.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution.* 38: 1358–1370.

Wilson PG, O'Brien MM, Heslewood MM, Quinn CJ. 2005. Relationships within Myrtaceae sensu lato based on a matK phylogeny. *Plant Systematics and Evolution.* 251:3-19.

## CUADROS

**Cuadro 1.** Sitios de colecta de material vegetal de “Cas” para determinar diversidad y estructura genética. **Población:** nombre de la población colectada, **PMA:** precipitación media anual (mm), **Temp:** temperatura anual. **Región:** regiones en donde se encuentran ubicadas las poblaciones cultivadas.

Población	Coordenadas geográficas		Elevación (msnm)	PMA (mm)	Temp (°C)	Individuos muestreados	Región
Cervantes	09°53'28.3"N	83°47'24.0"W	657	2500	24	15	Valle Central
Tacacorí	10°03'07.3"N	84°12'52.3"W	845	2100	23	15	
Ciruelas	09°59'05.7"N	84°15'26.3"W	810	1900	23	15	
Escazú	09°55'08.3"N	84°07'42.6"W	1300	1929	24	15	
Batán	10°04'34.4"N	83°22'37.2"W	114	3567	28	15	Caribe

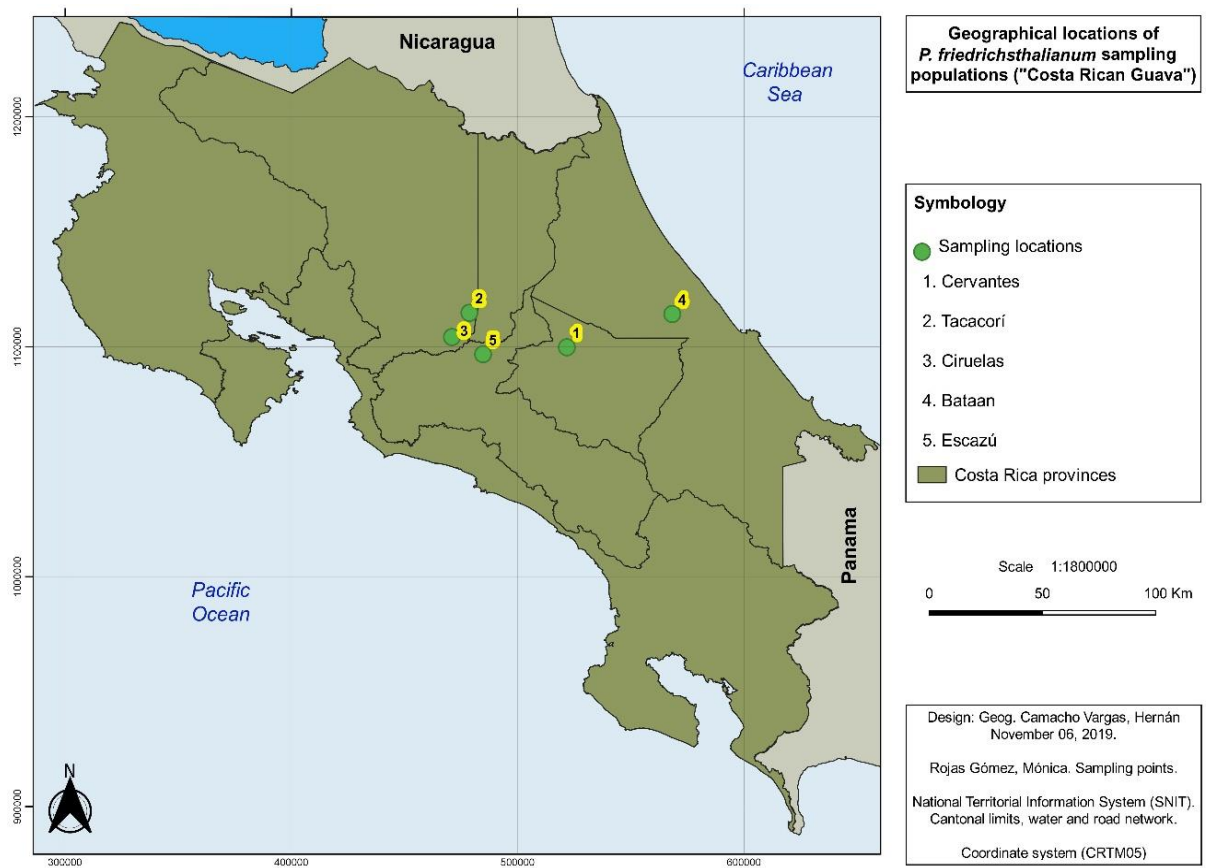
**Cuadro 2.** Estimativas básicas de diversidad genética de poblaciones cultivadas de “Cas” obtenidas en el presente estudio. **Ho**= heterocigosidad observada, **He**= heterocigosidad esperada, **Fis**=coeficiente de endogamia, **Fst**= estadístico de estructura genética. **DS**= desviación estándar

Populations	Ho	DS	He	DS	Alelos privados	Fst	Fis
Tacacorí	0.487	0.004	0.347	0.003	13	0.220	-0.412
Escazú	0.494	0.003	0.344	0.004	19		-0.445
Batán	0.485	0.005	0.349	0.005	15		-0.412
Ciruelas	0.479	0.002	0.347	0.002	15		-0.404
Cervantes	0.502	0.004	0.343	0.003	3		-0.457
Promedio	0.475	0.003	0.352	0.003	13		-0.426

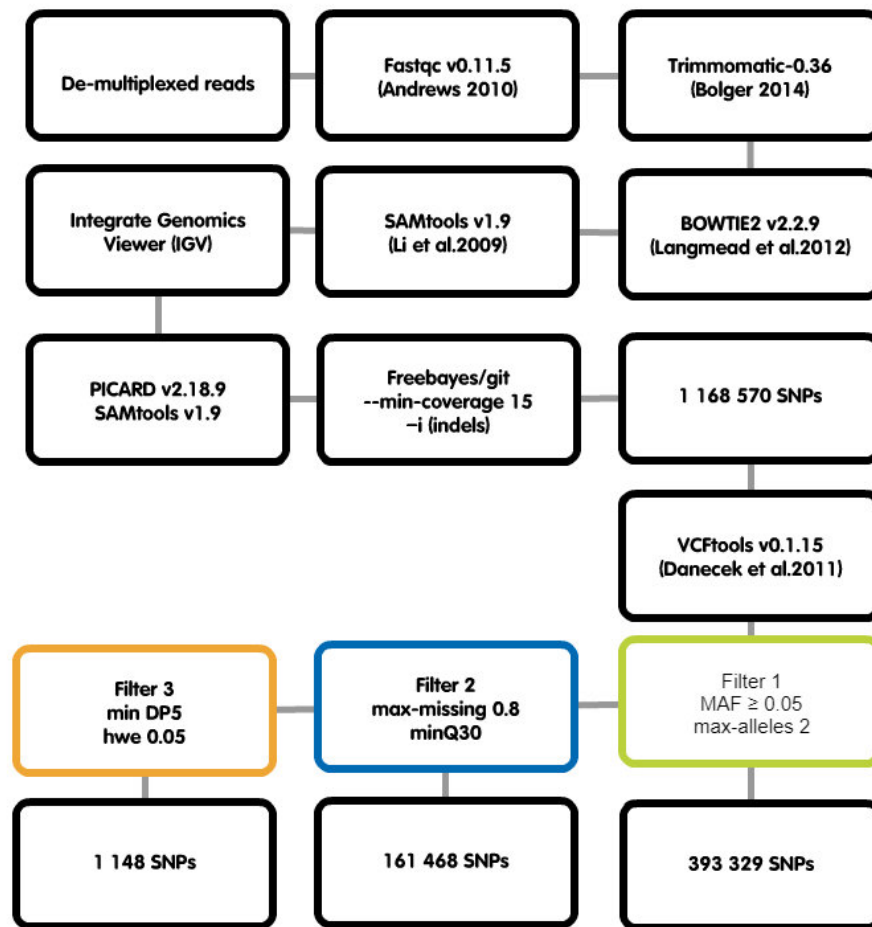
**Cuadro 3.** Distancias genéticas de Nei entre las poblaciones de “Cas” analizadas en el presente estudio.

	Escazú	Tacacorí	Batán	Ciruelas
Tacacorí	0.0063			
Batán	0.0075	0.0053		
Ciruelas	0.0084	0.0001	0.0047	
Cervantes	0.0052	0.0036	0.0046	0.0046

## FIGURAS

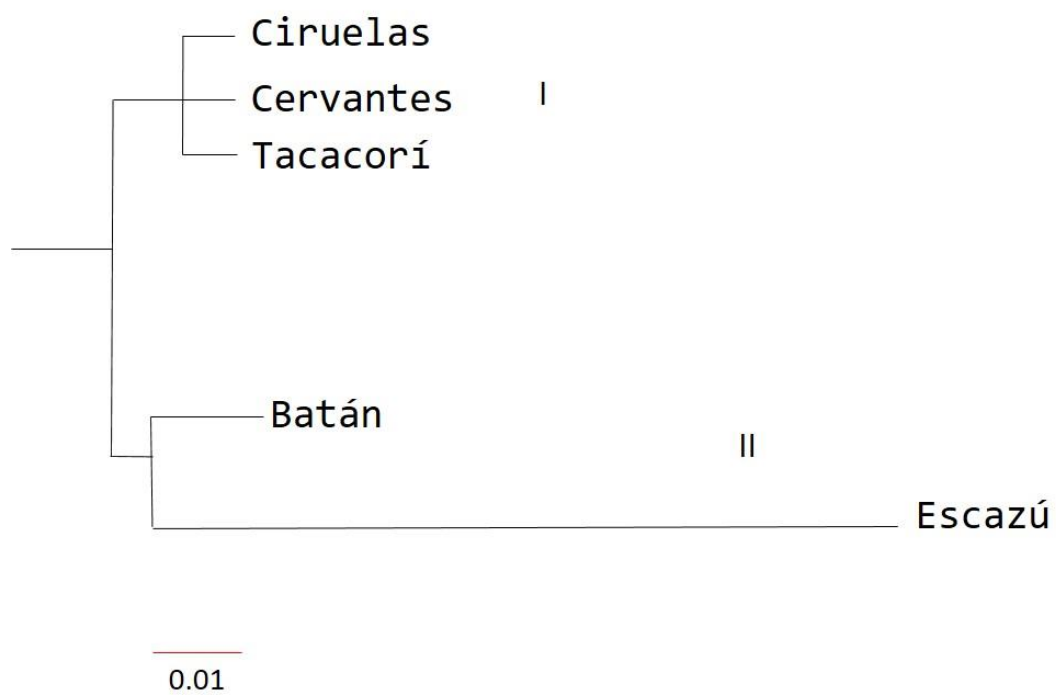


**Figura 1.** Mapa de ubicaciones geográficas donde se colectó el material vegetal de “Cas”. Los puntos marcados son puntos exactos del muestreo.

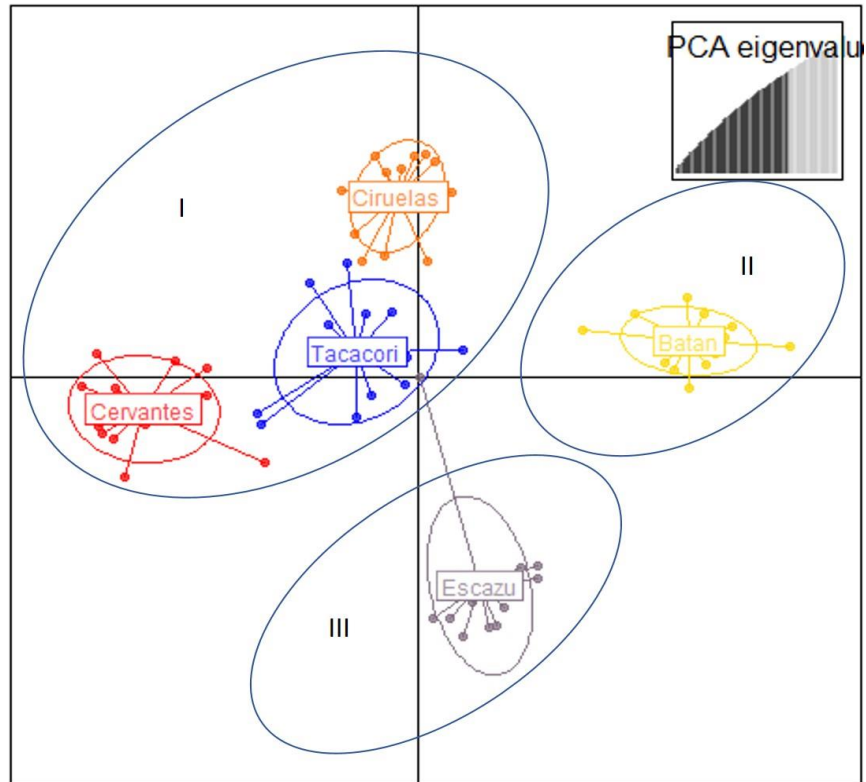


**Figura 2.** Pipeline bioinformático desarrollado para el análisis de las lecturas obtenidas por ddRAD-seq.

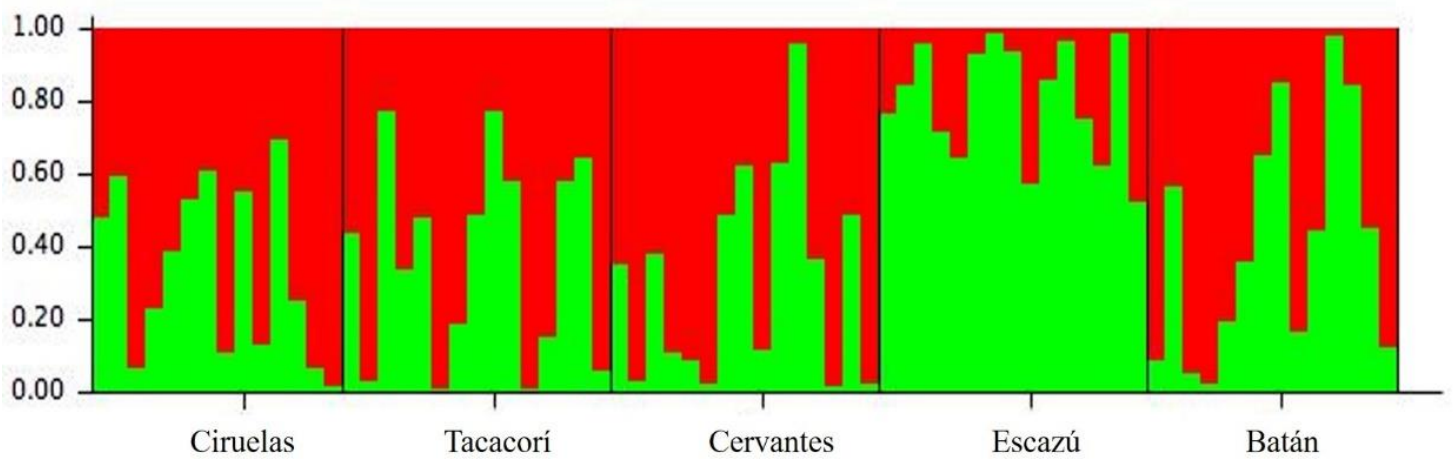




**Figura 3.** Dendrograma basado en las distancias genéticas de Nei utilizando el algoritmo de Neighbor-Joining (NJ) entre poblaciones de “Cas”.



**Figura 4.** Resultados del análisis discriminante de componentes principales (DAPC) basado en los datos de SNPs de “Cas”. Los ejes representan los dos primeros discriminadores lineales (LD); cada círculo representa un grupo y cada punto representa un individuo. Los números romanos y eclipses (de color azul) representan las diferentes subpoblaciones identificadas por el análisis DAPC.



**Figura 5.** Estructura de cinco poblaciones de “guayaba costarricense” revelada por análisis bayesiano (STRUCTURE), mostrando el mejor conglomerado de  $K=2$ . Cada columna vertical representa un individuo y diferentes colores en el mismo individuo indican el porcentaje del genoma compartido con cada grupo, de acuerdo con las proporciones de la mezcla. El eje y denota la probabilidad de la asignación de un individuo en las poblaciones muestreadas.

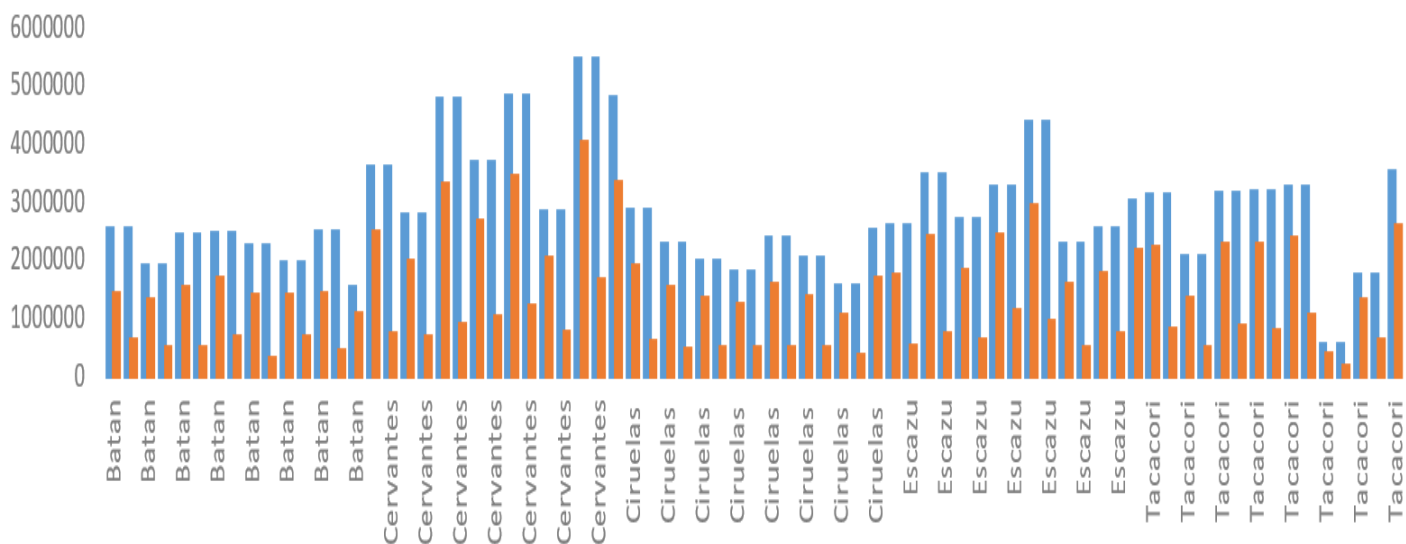
## MATERIAL SUPLEMENTARIO



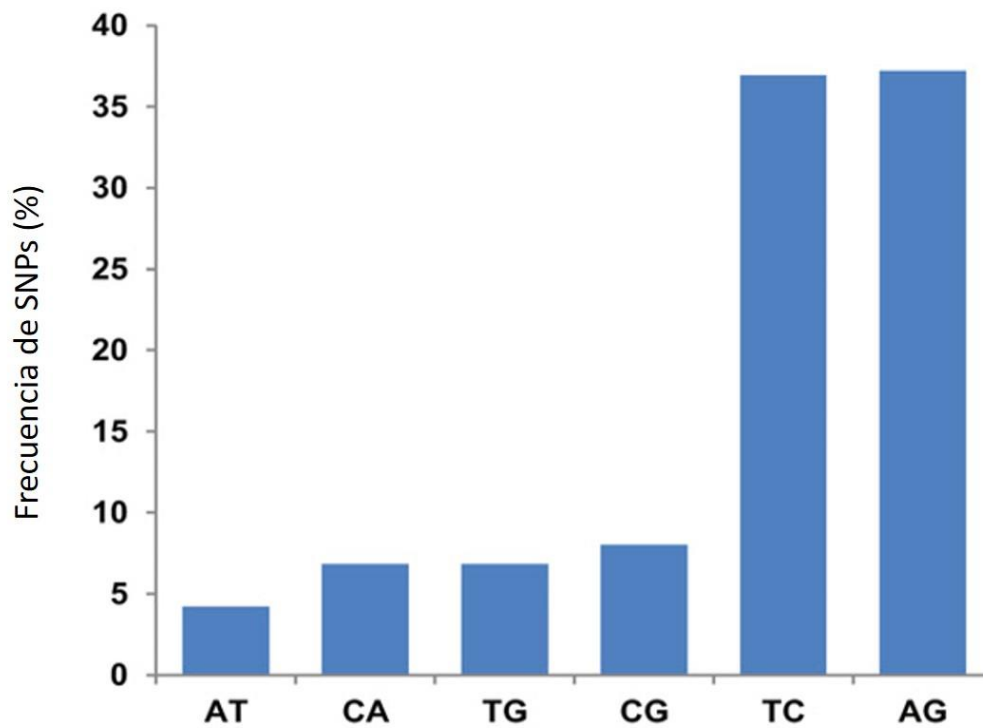
**Figura S1.** Imágenes representativas de algunas plantaciones de “guayaba costarricense” en donde se realizó la colecta del material vegetal para el presente estudio.



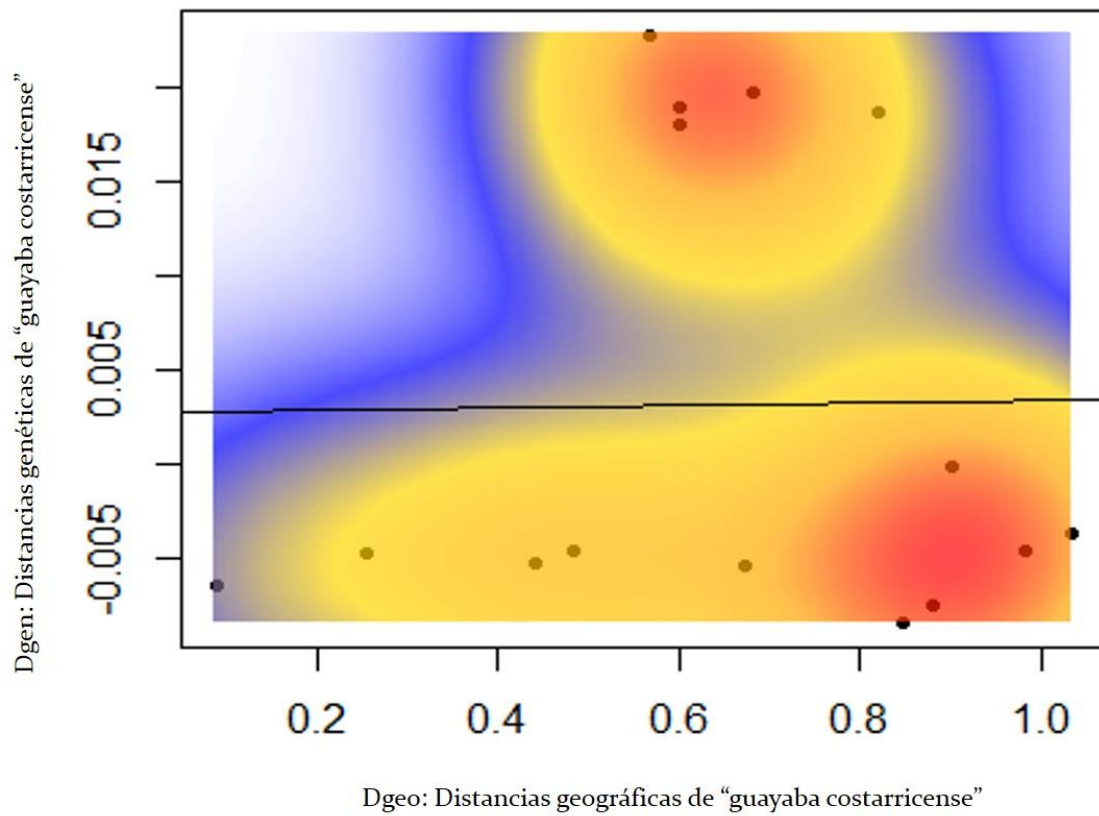
**Figura S2.** Alineamiento visualizado en IGV 2.3.x release de secuencias de algunos individuos de “guayaba costarricense” contra el congénere *Psidium guajava*.



**Figura S3.** Gráfica mostrando el número de lecturas antes (columnas azules) y después (columnas naranjas) del procesamiento con Trimmomatic-0.36 (SLIDINGWINDOW:4:30 y MINLEN:100) para cada individuo de las poblaciones de “guayaba costarricense” (*P. friedrichsthalianum*).



**Figura S4.** Porcentaje de transiciones ( $C \leftrightarrow T$  o  $A \leftrightarrow G$ ) (57.46%) y transversiones ( $A \leftrightarrow C$ ,  $C \leftrightarrow G$ ,  $A \leftrightarrow T$ ,  $G \leftrightarrow T$ ) (42.54%) observadas en el presente estudio.



**Figura S5.** Correlación de la distancia geográfica (Dgeo) y la distancia genética (Dgen) entre 75 individuos de poblaciones de "guayaba costarricense" (*P. friedrichsthalianum*), incluida la línea de regresión (prueba de Mantel, valor de correlación = 0.453, p-value = 0.097, en la aleatorización 999).



## Conclusiones generales

El objetivo del presente estudio fue caracterizar diversidad genética de poblaciones cultivadas de Cas (*Psidium friedrichsthalianum* (O. Berg) Nied) en Costa Rica, utilizando dos descriptores citogenéticos, número de cromosomas (ploidía) y tamaño del genoma; y un descriptor molecular, marcadores moleculares tipo SNPs obtenidos con ddRAD-seq. Esto con el fin de identificar germoplasma promisorio para su eventual conservación y/o selección en programas de mejoramiento genético.

En el primer capítulo observamos que las poblaciones de Cas en Costa Rica son consistentemente diploides ( $2n = 2x = 22$ ). Así mismo, determinamos que los individuos de dichas poblaciones presentan un contenido promedio de ADN nuclear de  $2C$  de  $1.960 \pm 0.005$  pg, equivalente a 1916.88 Mbp. La uniformidad encontrada entre las poblaciones costarricenses de Cas en términos de número cromosómico y tamaño del genoma (ADN nuclear) sugiere un origen común entre ellas. Alternativamente, sugerimos que dicha uniformidad puede ser una consecuencia del establecimiento de las plantaciones a partir de la selección artificial de individuos con rasgos de interés específicos y homogéneos, por ejemplo, tamaño de la fruta o alto contenido de pulpa.

Nuestros resultados sobre el número de cromosomas y el tamaño del genoma de la especie de Cas contribuyen a la caracterización citogenética de esta especie frutal importante como potencial alimento funcional, lo cual hasta el momento no era clara la información ni actualizada. Esta información puede usarse para diseñar estrategias de conservación regional que preserven los recursos genéticos locales. Así mismo, esta información podría ayudar a la definición taxonómica de *P. friedrichsthalianum* (Cas) y la comprensión de las relaciones filogenéticas entre otros miembros del género *Psidium*.

En el segundo capítulo, los marcadores moleculares tipo SNPs evidenciaron alta diversidad genética en las poblaciones cultivadas de Cas en Costa Rica; sugerimos que

esto puede deberse a la amplia distribución del Cas en el país, al sistema de reproducción cruzada que presenta el Cas, y la existencia de poblaciones naturales que mantienen el acervo genético original de la especie (Loveless & Hamrick 1984; Hamrick & Godt 1996). De la misma manera, es esperable observar alta diversidad genética en Cas debido a que este cultivo no ha pasado por procesos de domesticación más extensos, en comparación con otros cultivos (Ji *et al.* 2013). Por otra parte, las diferencias de elevación entre las poblaciones de Cas, el trasiego de material entre fincas aledañas y la forma en que los agricultores cultivan el Cas; parecen ser factores importantes de estructuración genética en estas poblaciones. De la misma manera, proponemos que el exceso de heterocigotas observado puede deberse principalmente al establecimiento de las plantaciones de Cas a partir de individuos o variedades altamente heterocigotas.

Finalmente, el material genético altamente diverso observado en el presente estudio es promisorio para su resguardo o conservación; dicho material puede utilizarse eventualmente para el establecimiento de programas de selección y mejoramiento genético del cultivo de Cas. Así mismo, es de importancia complementar dicha información con datos de perfiles de contenido de antioxidantes o con descriptores morfológicos (tamaño del fruto) de cada población analizada, para establecer la verdadera relación entre variantes genotípicas y fenotípicas, y seleccionar material genético altamente productivo con el cual se pueda establecer cultivos comerciales con un valor agregado de alimento funcional.